Ph.D. DISSERTATION

MAP BASED MOTION FIELD REFINEMENT METHODS FOR MOTION-COMPENSATED FRAME INTERPOLATION

움직임 보상 프레임 보간을 위한 최대 사후 확률 기반의 움직임 벡터장 교정 방법

BY

DOOSEOP CHOI

FEBRUARY 2014

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Abstract

In this dissertation, maximum *a posteriori* probability (MAP) based motion refinement methods are proposed for block-based motion-compensated frame interpolation (MCFI). The first method, called *single hypothesis Bayesian approach* (SHBA), is aiming at estimating the true MVF of a video frame from its observed MVF, which is the result of a block-based motion estimation (BME), by maximizing the posterior probability of the true MVF. For the estimation, the observed MVF is assumed to be a degraded version of the true MVF by locally stationary additive Gaussian noise (AGN), so the variance of the noise represents the unreliability of the observed MV. The noise variance is directly estimated from the observation vector and its select neighbors. The prior distribution of the true MVF is designed to rely on the distances between the MV and its neighbors and to properly smooth false MVs in the observation.

The second algorithm, called *multiple hypotheses Bayesian approach* (MHBA), estimates the true MVF of a video frame from its multiple observations by maximizing the posterior probability of the true. The multiple observations, which are the results of a BME incorporating blocks of different sizes for matching, are assumed to be degraded versions of the true MVF by locally stationary AGN. The noise variances for the observations are first estimated independently and then adaptively adjusted by block-matching errors in order to solve motion boundary problem.

Finally, a method, called *single hypothesis Bayesian approach in a bidirectional framework* (SHBA-BF), that simultaneously estimates the true forward and backward MVFs of two consecutive frames from the observed forward and backward MVFs is proposed. The observed MVFs are assumed to be degraded versions of the corresponding true MVFs by locally stationary AGN. The true forward and backward MVFs are

assumed to follow the proposed joint prior distribution, which is designed such that it adaptively relies on not only the resemblance between spatially neighboring MVs but also the resemblance between the MV and its dual MV so the proposed simultaneous estimation can fully exploit duality of MVF.

Experimental results show that the proposed algorithms obtain better or comparable performances as compared to the state-of-the-art BME algorithms at much lower computational complexity.

keywords: MAP-MRF framework, Motion-compensated frame interpolation (MCFI), Motion refinement, True motion estimation

student number: 2008-30246

Contents

List of Figures

List of Tables

Chapter 1

Introudction

1.1 Motion-Compensated Frame Interpolation

Motion-compensated frame interpolation (MCFI) has long been of great interest of researches concerning practical applications such as frame rate up conversion (FRUC) [23], image de-noising [1], and video compression [11]. It usually consists of two major steps: motion estimation (ME) and frame interpolation (FI). In the ME step, the motion vectors (MVs) of blocks or pixels of a video frame are estimated in relation to its surrounding frames. Block-matching approaches have widely been used for ME because of its simplicity and accuracy. Figure 1.1 shows an example. In the figure, the arrows present the true forward and backward MVs of blocks in the previous and current frames, respectively. In the FI step, one or more frames are interpolated using the estimated true MVs and the surrounding frames as illustrated in figure 1.2. In general, a new frame between the previous and current frame is interpolated by projecting blocks in the previous or current frame onto the new frame along the estimated motion trajectories of the blocks.

MCFI that uses block-based motion estimation (BME) suffers from annoying blocky artifacts and ghost effects as shown in Figure 1.3-(a) and (c). Figure 1.3-(a) and (c)

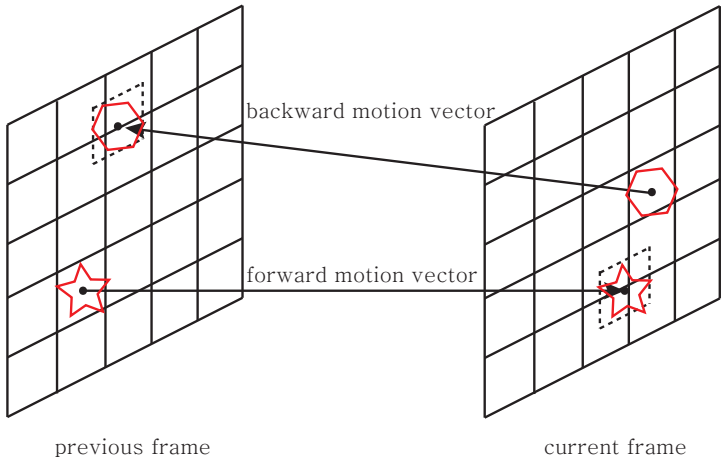


Figure 1.1: An example of block-based motion estimation.

respectively show a frame overlaid with the MVs estimated by a full search block-matching algorithm and the frame interpolated by the MVF in Figure 1.3-(a). This is because basic BME methods such as full search block-matching algorithm estimate MVs by only minimizing the block-matching errors and consequently produce false estimates. One of the solution to this problem is to produce smooth motion vector fields (MVFs) by taking into account the fact that the true MVFs of a common video are quite smooth except at motion boundaries. An example in Figure 1.3-(b) and (d) proves that smoothing MVF diminishes annoying blocky artifacts in the interpolated frame. Figure 1.3-(b) and (c) respectively show a frame overlaid with the MVs obtained by a MVF smoothing algorithm and the frame interpolated by the MVF in Figure 1.3-(b).

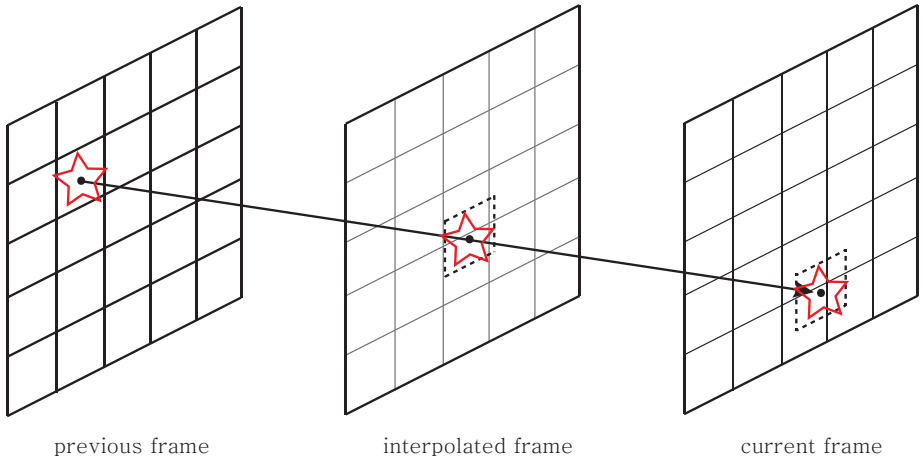


Figure 1.2: An example of frame interpolation.

1.2 Previous Works

In this section, the existing true MVF estimation algorithms are categorized according to how an algorithm produces smooth MVFs.

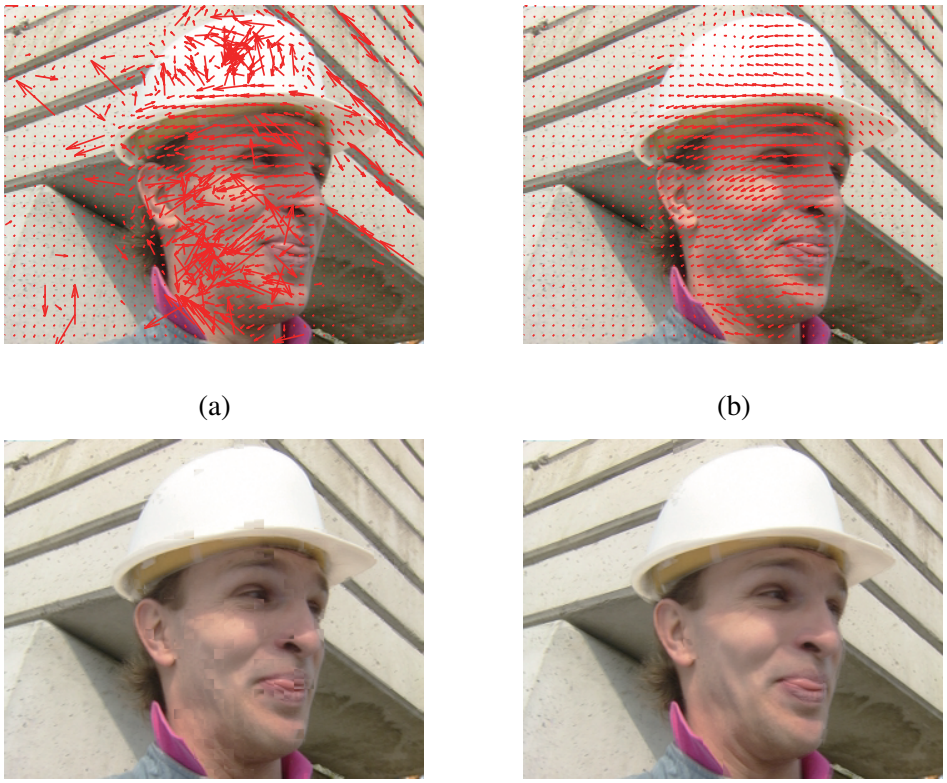
1.2.1 Exploit spatio-temporal correlation during ME

The methods in this category produce smooth MVFs by utilizing correlations between the spatio-temporally neighboring MVs. They usually estimate the true MVF by simultaneously minimizing the block- or pixel- matching errors and an irregularity measure of the MVF. Let I_p (the previous frame) and I_c (the current frame) denote two consecutive frames of a video. Then the true forward MVF of I_p is estimated by

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathcal{U}}{\operatorname{argmin}}\, \mathcal{E}(\mathbf{u}), \tag{1.1}$$

where the energy

$$\mathcal{E}(\mathbf{u}) = \sum_{i\in\mathcal{I}} \Big[\sum_{\mathbf{m}\in\mathcal{B}_i} |I_p(\mathbf{m}) - I_c(\mathbf{m}+\mathbf{u}(i))| + \lambda\phi(\nabla\mathbf{u}(i)) \Big], \tag{1.2}$$



(a) (b)

(c) (d)

Figure 1.3: Estimated MVFs ((a) and (b)) and frames interpolated using the MVFs ((c) and (d)).

$\mathcal{U}$ denotes the set of all candidate MVFs for I_p. $I_p(\mathbf{m})$ is the intensity value of the frame I_p at the pixel position $\mathbf{m}$, $\mathbf{u}(i)$ is a 2-D MV of the ith block of I_p, and $\mathcal{I}$ and $\mathcal{B}_i$ respectively denote the set of all block indices in a frame and the set of all pixel indices in the ith block. $\phi(\nabla\mathbf{u}(i))$, called *smoothness constraint*, is a regularization functional quantifying the smoothness of the MVF, and it is usually defined as $\phi(\nabla\mathbf{u}(i)) = \mathrm{trace}\{\nabla\mathbf{u}(i)^T\nabla\mathbf{u}(i)\}$ where ∇ denotes the vector gradient. λ is a constant, which trades off the block-matching errors against the smoothness constraint. The following summarizes the existing ME algorithms in this category.

Horn and Schunck [2] first proposed using a smoothness constraint in dense motion field estimation. The smoothness constraint is devised to give to the matching scores penalty in correlation with irregularity of the MVF. Nagel and Enkelmann [3], in the framework of Horm and Schunck, proposed an anisotropic smoothness constraint of the form $\phi(\nabla \mathbf{u}(i)) = \mathrm{trace}\{(\nabla \mathbf{u}(i))^T \mathbf{D}(\nabla \mathbf{u}(i))\}$ where the 2×2 diffusion matrix $\mathbf{D}$ makes use of the image gradient such that the smoothness constraint not simply smoothes MVFs but preserves motion boundaries. Konrad et al. [4] first introduced the dense motion field estimation method based on Bayesian inference. Because their prior distribution of the true MVF is defined so as to embody similarity between the neighboring MVs, their methods tends to produce smooth MVFs. They also introduced line fields to furnish the algorithm with a means to preserve motion boundaries. The same general goal let Stiller [5] propose an object-based, dense motion field estimation algorithm and let Chang [6] simultaneously estimate and segment the MVF under a Bayesian inference model. Black [7][8] and Wei [9], to preserve motion boundaries, employed discontinuity-adaptive smoothness constraints [10] in ME. The methods that simultaneously estimate the true MVFs (or optical flow fields or disparity fields) and occluded areas of successive frames of a video also have been proposed in the literature to preserve motion boundaries [11]-[19] based on the fact that occluded areas of a frame exist around motion boundaries and that the estimated MVs in the occluded areas usually show false motions. Hann et al. [20] estimated a true MV of the current block by choosing one from the set including the spatio-temporally neighboring MVs as well as and their perturbations. Wang et al. [21] proposed a multi-frame-based BME algorithm exploiting spatio-temporal correlation of MVFs through Bayesian inference. A recent research by Santoro et al. [22] incorporated the block-overlap cost, which penalizes the amount of overlaps between the blocks that are pointed by the MVs originating from non-overlapping blocks.

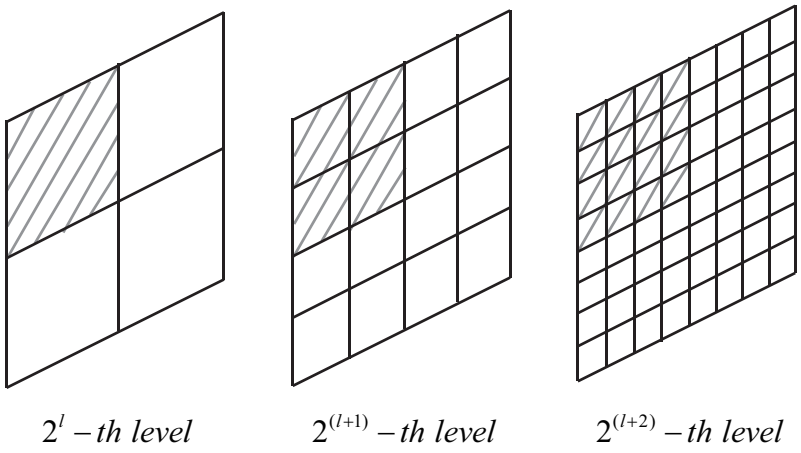


Figure 1.4: An example of block hierarchy in hierarchical ME.

1.2.2 Utilize multiple block sizes for matching in ME

The algorithms in this category, called hierarchical motion estimation (HME) algorithm, produce smooth MVFs by incorporating multiple block sizes for block-matching [28]-[33]. A reliable estimate of the true MV of a block can be obtained by increasing the size of the block for matching. This is because a larger block usually includes crucial features for the correct matching more than a smaller block. In contrary, the details of the true MVFs are well described as we decreases the size of the block for the matching. HME algorithms utilize the advantages of both approaches in order to produce smooth MVFs while preserve motion boundaries. Seo et al. [30] used the MV estimated on a larger block as an initial MV for the motion estimation on a smaller block. Choi et al. [28] proposed a MV reliability measure and used it to choose the most reliable one from the MVs estimated on matching blocks of different sizes. A work by Liu et al. [31] included a progressively reduced block-size motion estimation algorithm in their multiple hypotheses Bayesian frame interpolation method. They incorporated into the block-matching score a measure of MVF irregularity, which used not only the spatially neighboring MVs but also the one estimated on a larger matching

block. Recently, Kim et al. [33] proposed an iterative ME algorithm, in which a larger block is used for generating reliable MVs while a smaller block is used for correcting false MVs obtained by the larger block in an iterative manner.

1.2.3 Correct false motion vectors in given MVFs

Another approach is to detach smoothing and correcting issue from ME and "refine" each MV in relation to its neighboring vectors. The key topics in this category are how to measure the reliability of an estimated MV and how to correct an unreliable one using its neighboring MVs. The algorithms in this category are usually called *motion refinement* (MR) algorithm. A widely used method is the vector median filtering [34], which, though very simple, well smoothes MVFs. Alparone et al. [35] proposed an error-weighted VMF, in which the block-matching errors are used to calculate the filter weights. Because false motions across motion boundaries generally bring about large matching errors, the algorithm can deal with the problem of oversmoothing motion boundaries. In calculating filter weights, Sekiguchi et al. [36] incorporated, in addition to the matching errors, the variance of the neighboring MVs. Anandan [37] first introduced the iterative regularization framework to the MVF smoothing. The method proposed in [37] calculates the $k$the estimate of the true MVF by taking a weighted average of the observed MVF and the previous estimate, and the weights are devised to reflect the sharpness of the matching error surface at the minimum point. As the sharpness at the minimum point is in general linked to the dependability of the MV that points to the spot, the algorithm can avoid over-smoothing as well as correct false MVs. Singh [38] also proposed a regularization framework for MVF smoothing. The major difference is in that it utilizes, in addition to the sharpness of the matching error surface, the variance of the previously estimated MVF as a measure of the MVF's unreliability. An error-weighting method was incorporated by Zheng [39], as an anisotropic smoothness constraint, into the regularization method of [37].

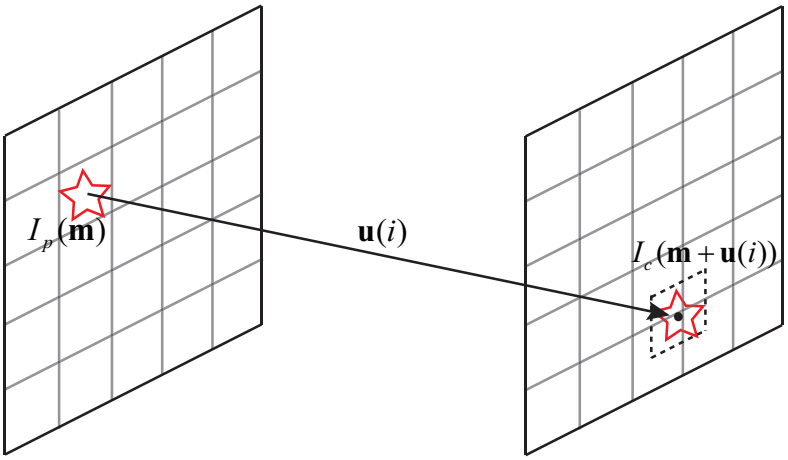


Figure 1.5: Image intensity consistency assumption.

Since the matching errors of the blocks or pixels in an occluded area are usually large, the smoothness constraint is designed so that the algorithm pays little attention to the neighboring MVs with large matching errors to avoid blurring motion boundaries. Sohn et al. [40] focused the regularization algorithm on dealing with false MVs in areas with spatially periodic patterns, also using the matching errors and the variance of the MVF. The same problem was addressed in the work by Lee et al. [41], where, by analyzing the shape of the matching error surface, false MVs are first detected and then replaced by the average of the neighboring MVs. Huang et al. [42][43] proposed merging blocks having unreliable MVs into a larger one on which they assigned on of the neighboring MVs based on the minimum matching error criterion. Dane et al. [44], to correct false MVs, proposed an adaptive use of the VMF based on the angle variance of the neighboring MV. Recently Jacobson et al. [45] proposed carefully refining MVs of salient moving objects in consecutive frames by a multistage MV refinement process while assigning a consistent motion to the backgrounds.

1.3 Motivation

MAP-MRF based ME algorithms estimate the true MVF of a frame by maximizing the posterior probability $p(\mathbf{u}|I_p, I_c)$ of the true MVF as follow.

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathcal{U}}{\operatorname{argmax}}\, p(\mathbf{u}|I_p, I_c) = \underset{\mathbf{u}\in\mathcal{U}}{\operatorname{argmax}}\, p(I_c|\mathbf{u}, I_p)p(\mathbf{u}|I_p), \tag{1.3}$$

where the second equation is due to Bayes rule. Image intensity consistency (IIC) assumption has been utilized for the observation model in order to define the observation likelihood $p(I_c|\mathbf{u}, I_p)$ in (1.3). Figure 1.5 describes IIC assumption. It means that if $\mathbf{u}(i)$ is the true MV of the $\mathbf{m}$th pixel of the frame I_p, the motion-compensated (MC) matching error $I_p(\mathbf{m}) - I_c(\mathbf{m} + \mathbf{u}(i)) = \epsilon$ is close to zero. If the MC matching errors are assumed to be realizations of identical random variables following Laplacian distribution and independent each other, the observation likelihood is positively correlated to $\exp(-\sigma \cdot |\epsilon|)$ [48], therefore, the likelihood can be expressed in terms of the block-matching errors. The prior distribution $p(\mathbf{u}|I_p)$ of the true MVF is usually expressed in terms of the irregularity measure of the MVFs such that $p(\mathbf{u}|I_p) \propto \exp(-\phi(\nabla\mathbf{u}(i)))$. Consequently, the problem of maximizing the posterior probability in (1.3) becomes that of minimizing the energy function in (1.2).

In this dissertation, Bayesian inference in the MAP-MRF framework is first employed for motion refinement (MR). Since MR is an algorithm aiming at correcting false MVs in *given* MVFs using the neighboring MVs, following three cases are considered in this dissertation.

Case 1

Given an observation $\mathbf{d}$ of the true MVF of a frame I_c, which is the results of a BME, the true MVF is estimated from the observation by

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathcal{U}}{\operatorname{argmax}}\, p(\mathbf{u}|\mathbf{d}, I_p, I_c). \tag{1.4}$$

Case 2

Given multiple observations $\mathbf{d}_1, ..., \mathbf{d}_M$ of the true MVF of a frame I_c, which are the results of a BME incorporating multiple block sizes for matching, the true MVF is estimated from the multiple observations by

$$\hat{\mathbf{u}} = \underset{\mathbf{u} \in \mathcal{U}}{\operatorname{argmax}}\, p(\mathbf{u}|\mathbf{d}_1, ..., \mathbf{d}_M, I_p, I_c). \tag{1.5}$$

Case 3

Given $\mathbf{d}_b$ and $\mathbf{d}_f$, which respectively are the observations of the true backward MVF and the true forward MVF and are the results of a BME, the true MVFs are estimated from the observations by

$$\{\hat{\mathbf{u}}_b, \hat{\mathbf{u}}_f\} = \underset{\{\mathbf{u}_b, \mathbf{u}_f\} \subset \mathcal{U}}{\operatorname{argmax}}\, p(\mathbf{u}_b, \mathbf{u}_f|\mathbf{d}_b, \mathbf{d}_f, I_p, I_c). \tag{1.6}$$

The solutions of the problems in (1.4), (1.5), and (1.6) will be called *single hypothesis Bayesian approach* (SHBA), *multiple hypotheses Bayesian approach* (MHBA), and *single hypothesis Bayesian approach in a bidirectional framework* (SHBA-BF), respectively.

Chapter 2

Single Hypothesis Bayesian Approach

In this chapter, a new motion vector field (MVF) refinement algorithm for block-based motion-compensated frame interpolation (MCFI) is proposed. Under the assumption that an observed MVF, such as the result of a block-based motion estimation (BME), is a degraded version of the true MVF, the true is estimated from the observation by maximizing the posteriori probability of it.

2.1 Problem Formulation

Let the notations first be introduced. Given two adjacent frames I_c (the current frame) and I_p (the previous frame) of a video, let $\mathbf{d} = \{\mathbf{d}(i)|i \in \mathcal{I}\}$ denote the 2-D backward motion vector field (MVF), obtained by a block-based motion estimation (BME) method, where $\mathbf{d}(i) = [d_x(i)\ d_y(i)]^T\ (\in \mathbb{R}^2)$ denotes the motion vector (MV) of the ith block of the frame I_c pointing to a matching block in I_p and $\mathcal{I}$ is the set of all block indices of a frame. Let us assume that $\mathbf{d}$, called the *observed* MVF or simply the *observation*, is a degraded version of the true MVF $\mathbf{u} = \{\mathbf{u}(i)|i \in \mathcal{I}\}$, where $\mathbf{u}(i) = [u_x(i)\ u_y(i)]^T\ (\in \mathbb{R}^2)$ is the true MV of the ith block. The degradation process

is expressed as

$$\mathbf{d} = \varphi(\mathbf{u}) \tag{2.1}$$

for some operation $\varphi(\cdot)$. Under this formulation, the true MVF $\mathbf{u}$ is estimated from the observation $\mathbf{d}$ by maximizing the posterior probability

$$p(\mathbf{u} \mid \mathbf{d}, I_p, I_c) = \frac{p(\mathbf{d} \mid \mathbf{u}, I_p, I_c)p(\mathbf{u} \mid I_p, I_c)p(I_p, I_c)}{p(\mathbf{d}, I_p, I_c)}. \tag{2.2}$$

Therefore, the maximum *a posteriori* (MAP) estimate $\hat{\mathbf{u}}$ of $\mathbf{u}$ is obtained by

$$\hat{\mathbf{u}} = \underset{\mathbf{u}\in\mathcal{U}}{\operatorname{argmax}}\, p(\mathbf{d} \mid \mathbf{u}, I_p, I_c)p(\mathbf{u} \mid I_p, I_c), \tag{2.3}$$

where $\mathcal{U}$ denotes the set of all candidate MVFs for I_c. $p(\mathbf{d}, I_p, I_c)$ and $p(I_p, I_c)$ of (2.2) are ignored because they are constant with respect to the unknown.

2.1.1 Proposed observation likelihood

Let us first look into the statistical nature of observation errors. Figure 2.1-(c) shows the normalized histogram of x-axis components of the observation error vectors. The error vectors are obtained by subtracting the ground-truth from the observed MVF of the test frame in Figure 2.1-(a). The observation is obtained by full search block-matching with block size 8×8 pixels and search range +/-24 both in vertical and horizontal direction at integer pixel accuracy. It is seen in the figure that the histogram shows a peak at zero and steep drops from the peak. The distribution of the error vectors is usually modeled as a generalized Gaussian distribution [46][50].

In Figure 2.1-(b), blocks of a similar texture content are presented in the same color. The green, red, and blue correspond to the low, mid, and high texture content, respectively. The amount of texture is measured by the variance of the pixels in the block. Figure 2.1-(d) shows the normalized histograms of the error vectors color-coded for the respective regions. We can observe two facts. The first is that the texture content of a block is inversely correlated with the magnitude of the error vector. This

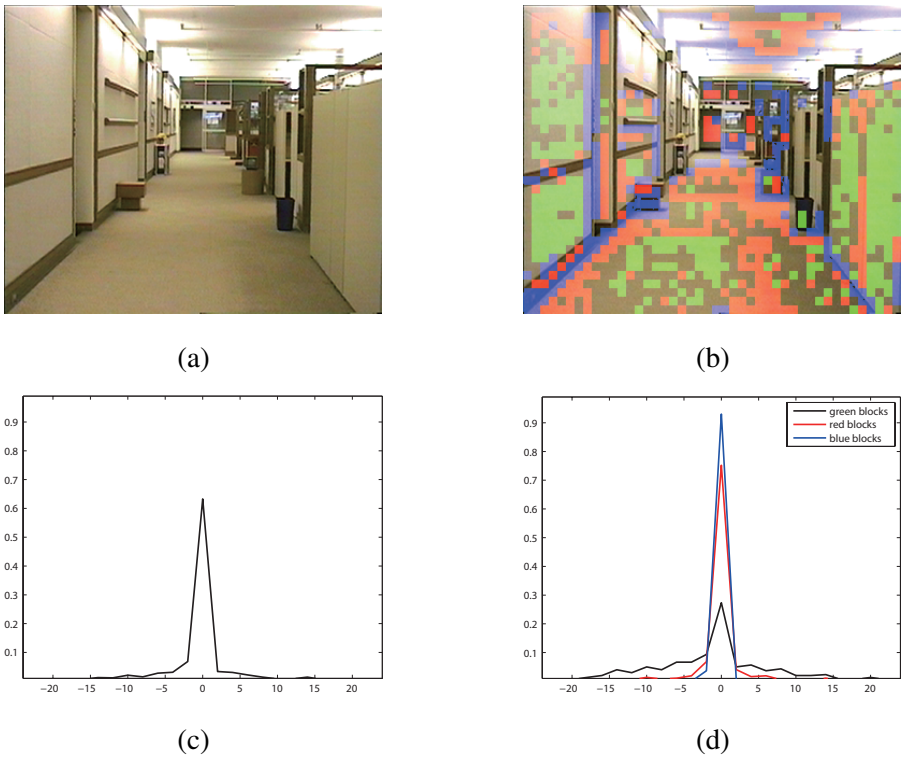


Figure 2.1: The normalized histograms of x-axis components of the observation error vectors. (a) test frame of *Hall*, (b) texture content classification; green, red, and blue correspond to low, mid, and high texture, respectively, (c) the normalized histogram of the error vectors for the full frame, (d) the normalized histograms of the error vectors color-coded for the respective regions.

is a natural result because finding a MV by block-matching relies heavily on the texture content. It can then be said that the statistics of the error vectors depend on the local image characteristics. The second is that blocks of the same color are clustered together. This implies that the error vector statistics may be considered not varying much in a small region of a frame. Therefore we can naturally consider that the error vectors assume a locally stationary distribution. It should be added that the same results are observed in the y-axis components of the error vectors and in other test

videos.

This nature of error vectors leads us to model the degradation operation $\varphi(\cdot)$ in (2.1) as locally stationary additive Gaussian noise such that

$$\mathbf{d}(i) = \mathbf{u}(i) + \mathbf{e}(i), \ \ \forall i \in \mathcal{I}, \tag{2.4}$$

where $\mathbf{e}(i)$ is a realization of a 2-D random noise vector, which is assumed Gaussian with the mean $\mathbf{0} = [0\ 0]^T$ and the diagonal covariance matrix

$$C_i = \begin{bmatrix} \eta_i & 0 \\ 0 & \kappa_i \end{bmatrix}. \tag{2.5}$$

Local stationarity of the AGN means that C_i may not stay the same value over the whole frame but may be considered the same over several adjoining blocks.

Two assumptions are made in calculating the observation likelihood $p(\mathbf{d} \mid \mathbf{u}, I_p, I_c)$. Assumed are that the observations $\mathbf{d}(i)$, $\forall i \in \mathcal{I}$, are conditionally independent given $\mathbf{u}$, I_p, and I_c, and that each $\mathbf{d}(i)$ is independent of $\mathbf{u}(j)$, $j \in \mathcal{I}$, $j \neq i$ given $\mathbf{u}(i)$, I_p and I_c, so that $p(\mathbf{d}(i)|\mathbf{u}, I_p, I_c) = p(\mathbf{d}(i)|\mathbf{u}(i), I_p, I_c)$. It then follows that $p(\mathbf{d} \mid \mathbf{u}, I_p, I_c) = \Pi_{i\in\mathcal{I}} p(\mathbf{d}(i) \mid \mathbf{u}(i), I_p, I_c)$. Since the additive noise model implies that

$$p(\mathbf{d}(i) \mid \mathbf{u}(i), I_p, I_c) = \frac{1}{2\pi\sqrt{\eta_i\kappa_i}} \exp\Big(-\frac{1}{2}[\mathbf{d}(i) - \mathbf{u}(i)]^T C_i^{-1} [\mathbf{d}(i) - \mathbf{u}(i)]\Big), \tag{2.6}$$

the observation likelihood $p(\mathbf{d}|\mathbf{u}, I_p, I_c)$ is obtained as follows:

$$p(\mathbf{d} \mid \mathbf{u}, I_p, I_c) = \Pi_{i\in\mathcal{I}} p(\mathbf{d}(i) \mid \mathbf{u}(i), I_p, I_c) = \\ \exp\Big(-N \ln 2\pi - \frac{1}{2}\sum_{i\in\mathcal{I}} \ln \eta_i\kappa_i - \sum_{i\in\mathcal{I}} \frac{1}{2}[\mathbf{d}(i) - \mathbf{u}(i)]^T C_i^{-1} [\mathbf{d}(i) - \mathbf{u}(i)]\Big) \tag{2.7}$$

Note that the diagonals of the covariance matrix in (2.5) represent the variance of the error between the true MV and the observation in horizontal and vertical directions. They are estimated from the observation, as will be described in section 2.2.

2.1.2 New prior distribution for true motion vector field

Many prior distributions have been proposed in the literature in order to model the characteristic of true MVFs of a common video, namely *piecewise smoothness*. As mentioned in Introduction, the image gradients have been incorporated in the distributions in order to describe motion boundaries of the true MVF, reflecting the fact that the intensity edges coincide with the motion boundaries. The image gradients, however, are not appropriate for describing motion boundaries of the block-grid MVFs since intensity edges does not always coincide with pixel block boundaries.

In this section, a new prior distribution of the true MVF utilizing block-matching errors to describe the motion boundaries is proposed. The matching errors also serve as a measure, in conjunction with the vector distances between the neighboring MVs, of how well a neighboring MV presents the true motion of the center block.

Terminology

Let $\epsilon_i(\mathbf{v})$ denote the sum of absolute differences (SAD) of the ith block corresponding to the MV $\mathbf{v}$, such that, with $I_c(\mathbf{m})$ indicating the intensity value of the grayscale representation of the frame I_c at the pixel index $\mathbf{m}$,

$$\epsilon_i(\mathbf{v}) = \sum_{\mathbf{m}\in\mathcal{B}_i} |I_c(\mathbf{m}) - I_p(\mathbf{m}+\mathbf{v})|, \tag{2.8}$$

where $\mathcal{B}_i$ denotes the set of all the pixel indices in the ith block. Let $\mathcal{N}_i$ denote the index set of the 8 blocks neighboring the ith block. To discriminate the neighboring blocks that are unlikely to be in the same moving object as the ith block, the subset $\mathcal{N}_i^{\mathbf{u}}$ of $\mathcal{N}_i$, called *suitable set*, is defined such that

$$\mathcal{N}_i^{\mathbf{u}} = \{j | j \in \mathcal{N}_i, x_i(j) = 1\}, \;\; i \in \mathcal{I}, \tag{2.9}$$

where the *suitability indicator* $x_i(j)$ is given by

$$x_i(j) = \begin{cases} 1, & \epsilon_i(\mathbf{u}(j)) \leq \lambda^{\mathbf{u}} \\ 0, & \epsilon_i(\mathbf{u}(j)) > \lambda^{\mathbf{u}} \end{cases} \tag{2.10}$$

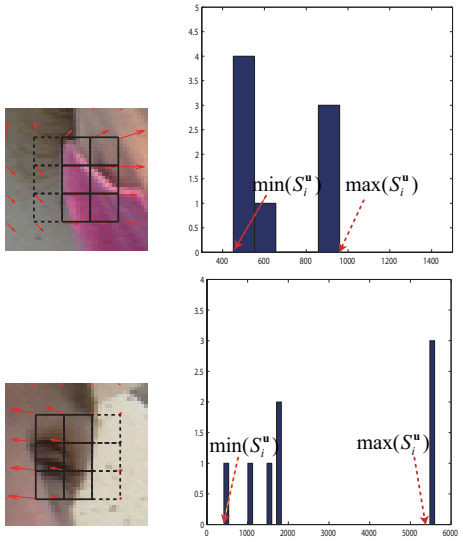


Figure 2.2: (a) 8×8 blocks around motion boundaries, (b) histogram of SADs for the center block when the ground-truth MVs of the neighboring blocks are used.

with a threshold $\lambda^{\mathbf{u}}$. The suitability indicator is based on the observation that the MVs of the neighboring blocks not belonging to the same moving object as the center block usually result in larger block-matching errors than those of the neighboring blocks belonging to the same moving object when applied to the center block. Figure 2.2 shows examples. Figure 2.2-(a) shows 8×8 blocks around motion boundaries and their ground-truth MVs, placed at top left corner of each block. The ground-truth MVs are obtained from full search block-matching results by manual selection with the help of VMF. Dashed lines indicate the blocks not belonging to the same moving object as the center block. Figure 2.2-(b) shows the histograms of the SAD values when the ground-truth MVs of the 8 neighboring blocks are applied to the center block. 128 is used for the width of the histogram bins. It is found from the figure that the MVs of the dashed blocks result in larger SAD values than those of the solid blocks.

For the suitability indicator to work right, the threshold $\lambda^{\mathbf{u}}$ in (2.10) needs to be

chosen properly. We need $\lambda^{\mathbf{u}}$ that picks out the minimum SAD of the center block $\min(\mathcal{S}_i^{\mathbf{u}})$, where $\mathcal{S}_i^{\mathbf{u}} = \{\epsilon_i(\mathbf{u}(j)) \mid j \in \mathcal{N}_i\}$, such that

$$\lambda^{\mathbf{u}} = \min(\mathcal{S}_i^{\mathbf{u}}) + \delta \tag{2.11}$$

for some δ. A correct value of δ, however, should depend on the local characteristics of the target frame, such as the luminance of the moving object and the background, and therefore can not be a constant throughout a video.

In order to deal with video sequences with diverse motions and scene characteristics in a frame, a variable δ is needed. In this dissertation, one defined by

$$\delta = \begin{cases} \max(\mathcal{S}_i^{\mathbf{u}}) - \min(\mathcal{S}_i^{\mathbf{u}}), & \text{if } \max(\mathcal{S}_i^{\mathbf{u}}) - \min(\mathcal{S}_i^{\mathbf{u}}) \leq \Delta \\ 0.5 \cdot (\max(\mathcal{S}_i^{\mathbf{u}}) - \min(\mathcal{S}_i^{\mathbf{u}})), & \text{otherwise} \end{cases} \tag{2.12}$$

is proposed where Δ is a constant. The following is a reason behind this choice. If a block and all its neighbors belong to the same moving object, their MVs, even if interchanged, usually give nearly the same SADs. This situation is identified by the condition $\max(\mathcal{S}_i^{\mathbf{u}}) - \min(\mathcal{S}_i^{\mathbf{u}}) \leq \Delta$ and needs $\lambda^{\mathbf{u}} = \max(\mathcal{S}_i^{\mathbf{u}})$ to include all the neighboring MVs in the suitable set, hence $\delta = \max(\mathcal{S}_i^{\mathbf{u}}) - \min(\mathcal{S}_i^{\mathbf{u}})$. The value of Δ needs to be reasonably small to strike a balance between false inclusion and false rejection. It is found from experiments that $\Delta = 2|\mathcal{B}_i|$ gives overall satisfactory results on test videos presented in chapter 5. On the other hand, if a block is located at a motion boundary, interchange of the neighboring MVs may give large SADs as shown in the histogram of Figure 2.2-(b). In this case, $\lambda^{\mathbf{u}}$ needs to be small enough to discriminate MVs of different moving objects. Figure 2.2 and further experiments on various test videos tell that δ for the correct discrimination usually turns out to be around $0.5 \cdot (\max(\mathcal{S}_i^{\mathbf{u}}) - \min(\mathcal{S}_i^{\mathbf{u}}))$.

Proposed prior distribution model

It is assumed that the true MVF of the frame I_c follows the distribution

$$p(\mathbf{u}|I_p, I_c) = \frac{1}{Z} \exp\Big(-\gamma \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}_i} V_{ij}(\mathbf{u}) \Big), \tag{2.13}$$

where $Z = \sum_{\mathbf{u} \in \mathcal{U}} \exp(-\gamma \sum_i \sum_j V_{ij}(\mathbf{u}))$ is a normalizing constant, $\mathcal{U}$ denotes the set of all candidate MVFs of I_c, and γ is a constant. The anisotropic potential function $V_{ij}(\mathbf{u})$ proposed in this dissertation has the following form:

$$V_{ij}(\mathbf{u}) = \mu_{ij} ||\mathbf{u}(i) - \mathbf{u}(j)||^2. \tag{2.14}$$

The probability $p(\mathbf{u}|I_p, I_c)$ is designed to increase when the neighboring MVs resemble each other more closely. The weight factor $0 \leq \mu_{ij} \leq 1$ in (2.14) controls the influence of the resemblance on the probability. This factor, as will be shown in section 2.3, consequently manages the influence of the neighboring MVs on the center MV $\mathbf{u}(i)$ during the refinement process. Therefore, μ_{ij} should be large if a neighbor $\mathbf{u}(j)$ is likely to reflect the true motion of the ith block. Otherwise, it should be small. These requirements lead us to define it as a combination

$$\mu_{ij} = (1 - \zeta)\alpha_{ij} + \zeta \beta_{ij}, \;\; 0 \leq \zeta \leq 1, \tag{2.15}$$

of the *likelihood measure* α_{ij} and the *dominance measure* β_{ij}, where

$$\alpha_{ij} = \frac{\epsilon_i(\mathbf{u}(j))^{-1}}{\sum_{k \in \mathcal{N}_i^{\mathbf{u}}} \epsilon_i(\mathbf{u}(k))^{-1}} \cdot x_i(j), \tag{2.16}$$

$$\beta_{ij} = \frac{[\sum_{l \in \mathcal{N}_i^{\mathbf{u}}} \|\mathbf{u}(j) - \mathbf{u}(l)\|^2]^{-1}}{\sum_{j \in \mathcal{N}_i^{\mathbf{u}}} [\sum_{l \in \mathcal{N}_i^{\mathbf{u}}} \|\mathbf{u}(j) - \mathbf{u}(l)\|^2]^{-1}} \cdot x_i(j). \tag{2.17}$$

Note that $0 \leq \mu_{ij} \leq 1$ and $\sum_{j \in \mathcal{N}_i} \mu_{ij} = 1$.

α_{ij} is introduced in order to make the center MV $\mathbf{u}(i)$ get influenced mainly by the neighboring MVs that minimize block-matching errors of the center block. This is because the true MV usually minimizes block-matching errors. It is called the likelihood measure because minimizing the block-matching error of the ith block is the

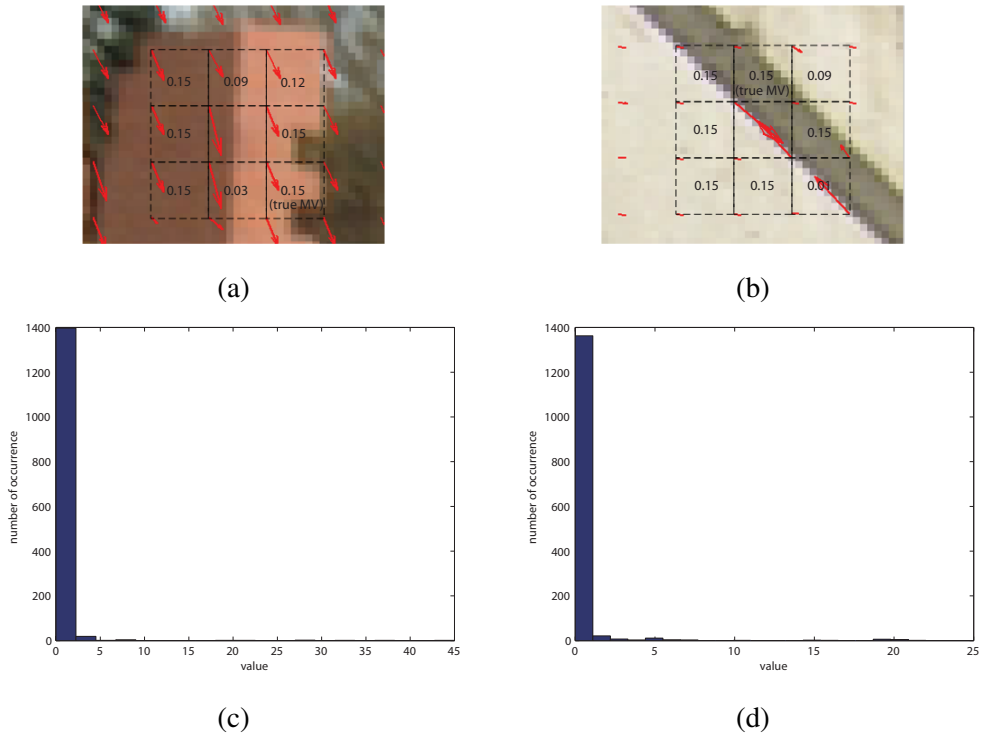


Figure 2.3: (a),(b) 8×8 pixel blocks and their MVs. The MVs are obtained from a full-search block matching algorithm. The scores shown are β_{ij} values. (c),(d) histograms of vector distances between the MV with the highest β_{ij} and the solution of (2.18). Note that (c) and (d) are obtained from the MVFs in Figure 2.8-(b) and (c), respectively.

same as maximizing the likelihood $p(I_p|\mathbf{u}(i) = \mathbf{v}, I_c)$, the probability that the ith block of I_c occurs at the location of I_p that $\mathbf{v}$ points to, under the assumption that the block-matching errors are independent each other and follow Laplace distribution [48]. The dominance measure β_{ij} counteracts the occasional false effect of α_{ij}. The false effect happens when an image has a spatially periodic pattern or is shift-invariant in a certain direction. Figure 2.3-(a) shows an example, in which the MV, placed at the top left corner of the center block, shows a false motion even if the MV results in small matching error. This is because the vertical edge of the image makes it nearly

shift-invariant. One of the solutions to this problem is to increase the size of the block for matching so that the block may include non-shift-invariant part as well. The dominance measure β_{ij} reflects this solution. By incorporating the differences between the MVs in the suitable set $\mathcal{N}_i^{\mathbf{u}}$ as seen in the definition, it gives a higher weight to a more dominant vector in the set. As a result, the contribution of a neighboring MV to the estimate of the true MV of the center block is smaller when it is less dominant in the set even if it gives a small SAD for the center block. The following is a probabilistic interpretation of how the dominance measure reflects the solution.

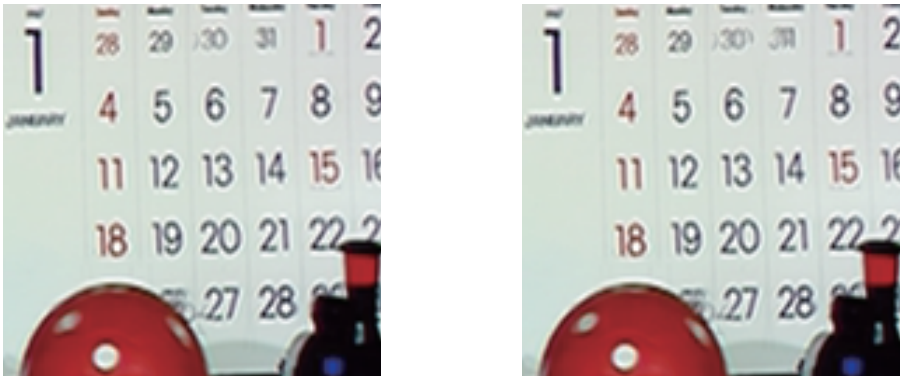
Suppose that the suitable set correctly identifies the blocks in the same moving object. Then, a reliable estimate of the true MV of the center block will be the one that minimizes the matching error for the union of the center block and its neighbors in the suitable set, that is,

$$\underset{\mathbf{v}}{\operatorname{argmin}}\, \epsilon_{\bar{\mathcal{B}}_i}(\mathbf{v}), \tag{2.18}$$

where $\bar{\mathcal{B}}_i = \{\mathbf{m} | \mathbf{m} \in \mathcal{B}_i \cup \{\mathcal{B}_j, j \in \mathcal{N}_i^{\mathbf{u}}\}\}$ ($\mathcal{B}_i$ denoting, as before, the set of pixel indices in the ith block) and $\epsilon_{\bar{\mathcal{B}}_i}(\mathbf{v}) = \sum_{\mathbf{m}\in\bar{\mathcal{B}}_i} |I_c(\mathbf{m}) - I_p(\mathbf{m}+\mathbf{v})|$. If we let $\mathbf{u}(\bar{\mathcal{B}}_i)$ denote the MV of the union of the blocks specified by $\bar{\mathcal{B}}_i$, a probabilistic equivalent of (2.18) would be $\operatorname{argmax}_{\mathbf{v}} p(I_p|\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}, I_c)$ since the conditional probability or the likelihood $p(I_p|\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}, I_c)$ is positively correlated to $\exp(-\sigma\epsilon_{\bar{\mathcal{B}}_i}(\mathbf{v}))$ if the matching errors for non-overlapping blocks are assumed independent of each other and follow Laplace distribution [48]. Bayes' rule and the assumption that $p(\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}, I_c)$ is a constant, i.e, that all candidate MVs are treated equally, gives us

$$\begin{aligned}
&\underset{\mathbf{v}}{\operatorname{argmax}}\, p(I_p|\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}, I_c) \\
&= \underset{\mathbf{v}}{\operatorname{argmax}}\, \frac{p(\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}|I_p, I_c)}{p(\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}, I_c)} \\
&= \underset{\mathbf{v}}{\operatorname{argmax}}\, p(\mathbf{u}(\bar{\mathcal{B}}_i) = \mathbf{v}|I_p, I_c).
\end{aligned} \tag{2.19}$$

It is argued by this development that finding the MV that minimizes the matching error for the pixels in $\bar{\mathcal{B}}_i$ is equivalent to finding the most probable MV of the same

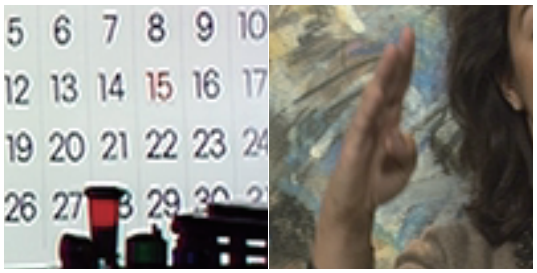


(a) $\zeta = 1.0$ (b) $\zeta = 0.0$

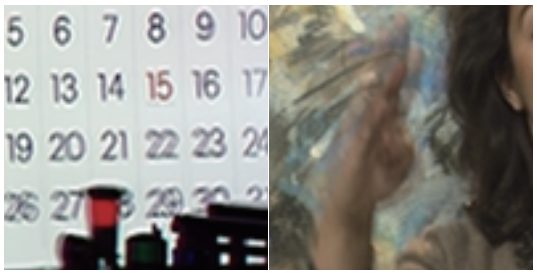
Figure 2.4: Frame interpolation results with different ζ.

region. Therefore, β_{ij}, which gives a higher weight to a more dominant vector in the suitable set, can be considered an alternative though approximate solution to (2.18). This is demonstrated by examples in Figure 2.3. In the figure, the suitable set comprises all 8 neighboring blocks, and the scores on the blocks are the calculated β_{ij} values. We can see that the most dominant MV among the suitable set, which is given the highest β_{ij} value, matches the true motion of the center block. From the experiments using the test videos in Figure 2.8, It is also found that the MV obtained by (2.18) coincides most of the times with the vector with the highest β_{ij} as seen in Figure 2.3-(c) and (d). The figures show the histograms of the vector distances between the MV with the highest β_{ij} value and the solution of (2.18).

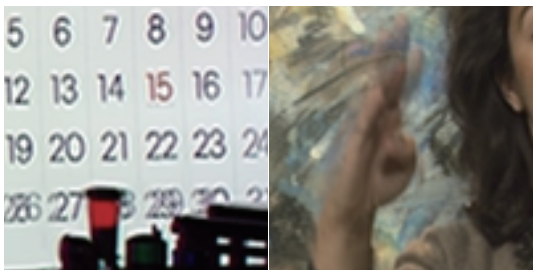
The parameter ζ, as mentioned above, trades off α_{ij} against β_{ij}. In order to show how ζ affects the MAP estimate of (2.3), the interpolation results of the 10-th frame of *Mobile* test video obtained from the 9-th and 11-th frame and the MAP estimates of their true backward MVF is presented in Figure 2.4. The red ball has a rotational motion while the background stands nearly still. We can see in the figure that the rotational motion of the ball is corrupted when $\zeta = 1$. (See the blurred white dot on the ball in Figure 2.4-(a).) This is because some of the MVs in the neighborhood do



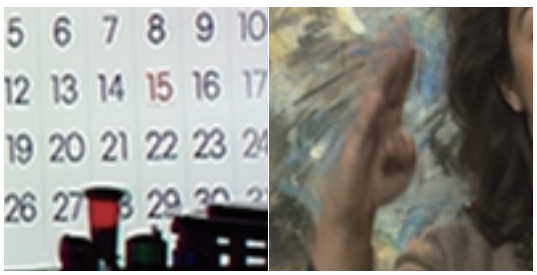
(a)



(b)



(c)



(d)

Figure 2.5: Performance evaluation of the proposed potential function. (a) original frame, (b)$\sim$(d) interpolated frames using MAP estimates obtained under (b) $V_{ij}(\mathbf{u}) = \frac{1}{|\mathcal{N}_i|}||\mathbf{u}(i)-\mathbf{u}(j)||^2$, (c) $V_{ij}(\mathbf{u}) = \frac{1}{|\mathcal{N}_i|}||\mathbf{u}(i)-\mathbf{u}(j)||_1$, (d) $V_{ij}(\mathbf{u}) = \mu_{ij}||\mathbf{u}(i)-\mathbf{u}(j)||^2$.

not reflect the true motion of the center block while in the same moving object. On the other hand, $\zeta = 0$ fails to correct false MVs among a periodic pattern due to the false effect of α_{ij}. (See the numbers on the background in Figure 2.4-(b).) The figures indicate that ζ should depend on the scene characteristics of the target frame in order to obtain correct estimates. However, it is not easy to identify the characteristics prior to the estimation. So $\zeta = 0.5$ is used in the experiments in order to deal with diverse motions and scene characteristics.

The distribution of (2.13) is further discussed in section 2.5 to reveal the properties of the corresponding random vector fields. In such a vector field, the probability of a random vector depends not only on its neighbors but also on the neighbors of the neighbors. This dependence might be termed *semi-markovianity*. This property, like Markovianity, encourages neighborhood dependency, which is a prime factor in smooth MVF models.

Finally Figure 2.5 shows how well the proposed potential function works for correcting false MVs in the observed MVF during the estimation compared to existing potential functions. Figure 2.5-(a) shows the original frames while Figure 2.5-(b)$\sim$(d) show the frame interpolated using MAP estimates. In order to obtain Figure 2.5-(b)$\sim$(d), the MAP estimates are calculated using $V_{ij}(\mathbf{u}) = \frac{1}{|\mathcal{N}_i|}||\mathbf{u}(i) - \mathbf{u}(j)||^2$, $V_{ij}(\mathbf{u}) = \frac{1}{|\mathcal{N}_i|}||\mathbf{u}(i) - \mathbf{u}(j)||_1$, and $V_{ij}(\mathbf{u}) = \mu_{ij}||\mathbf{u}(i) - \mathbf{u}(j)||^2$, respectively. We can see in the figure that the proposed potential function corrects false MVs better than the existing potential function both at motion boundaries and on images with periodic patterned areas.

2.2 Estimation of AGN Variance

The noise covariance matrix C_i, in improving the estimate of the true MV of the ith block, trades off the contribution of the observation against that of the neighboring MVs, as will be seen in section 2.3. The trade-off should naturally depend on the

confidence we have on the observation [3][37]. For example, if the observation seems reliable, it should be given higher weight, thus avoiding over-smoothing. Otherwise, it should be given lower weight so that it gets well smoothed out by the neighboring MVs.

Many researches have been done in order to measure the reliability of the estimated MV. Widely used measures are the minimum SAD value [21][42][43], the sharpness of the matching error surface at its minimum point [37][39], and the variance of the MVF [21][43]. Patras et al., in [48], argued that the reliability of the MV obtained from a BME algorithm is highly correlated with the prominence of the minimum peak of the matching error surface. Recently, Santoro et al. [22] included in the measure the amount of overlaps between the blocks that are pointed to by the motion vectors originating from non-overlapping blocks.

In this section, a method that estimates the covariance matrix of locally stationary additive Gaussian noise (AGN) directly from the observation is presented. The covariance matrix estimated by the proposed method shows strong correlation with the reliability of the observation since it reflects the prominence of the minimum peak of the matching error surface.

2.2.1 Proposed covariance matrix estimation method

The covariance matrix C_i is estimated through maximum likelihood estimation. Accurate estimation requires a sufficient number of the observations of $\mathbf{u}(i)$, while we are given only one observation $\mathbf{d}(i)$. To overcome this problem, the observed MVs $\{\mathbf{d}(l)|l \in \mathcal{N}_i\}$ in the neighborhood can be regarded as additional observations of $\mathbf{u}(i)$.

As we observed in section 2.1.1, the variance of the observation error vectors depends on the local image characteristics such as the amount of textures. So the neighboring observation MVs to be included in estimation of C_i should be those that better reflect the image characteristic of the ith block. Experiments tell us that the MVs

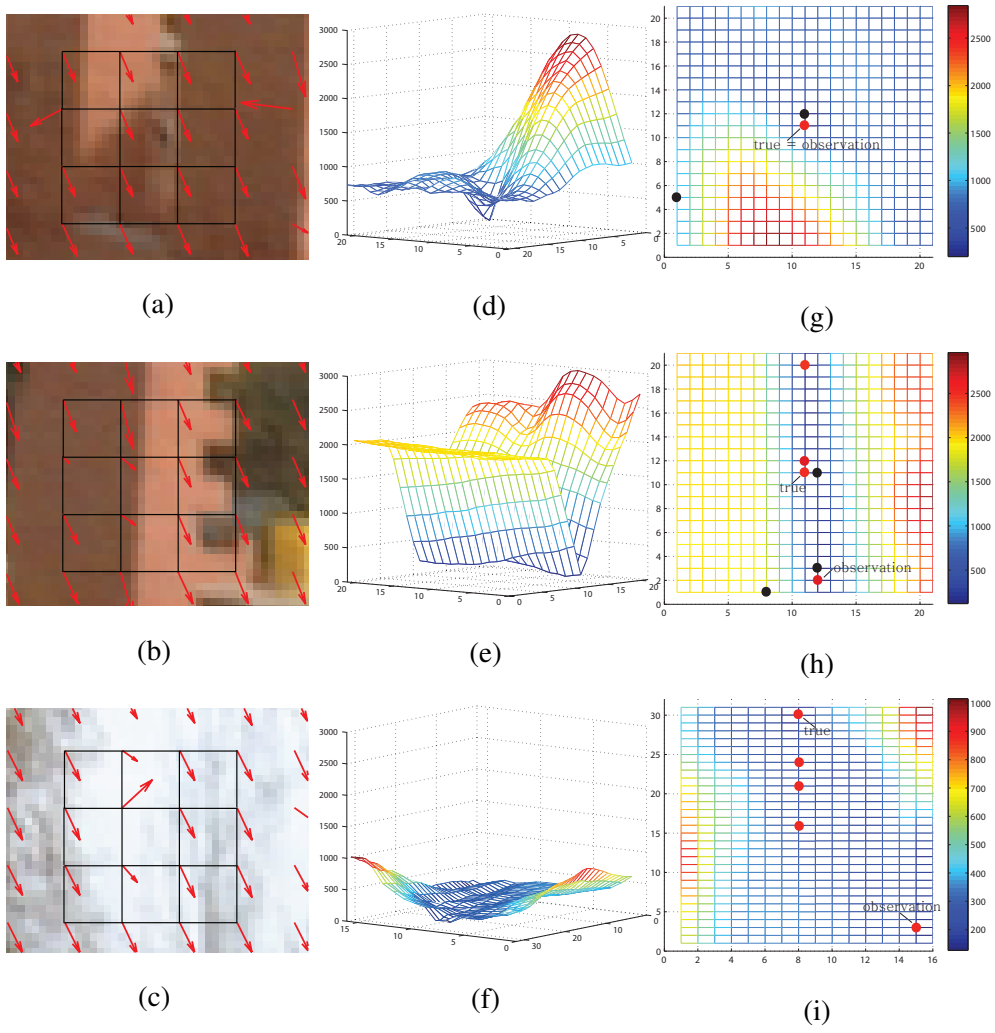


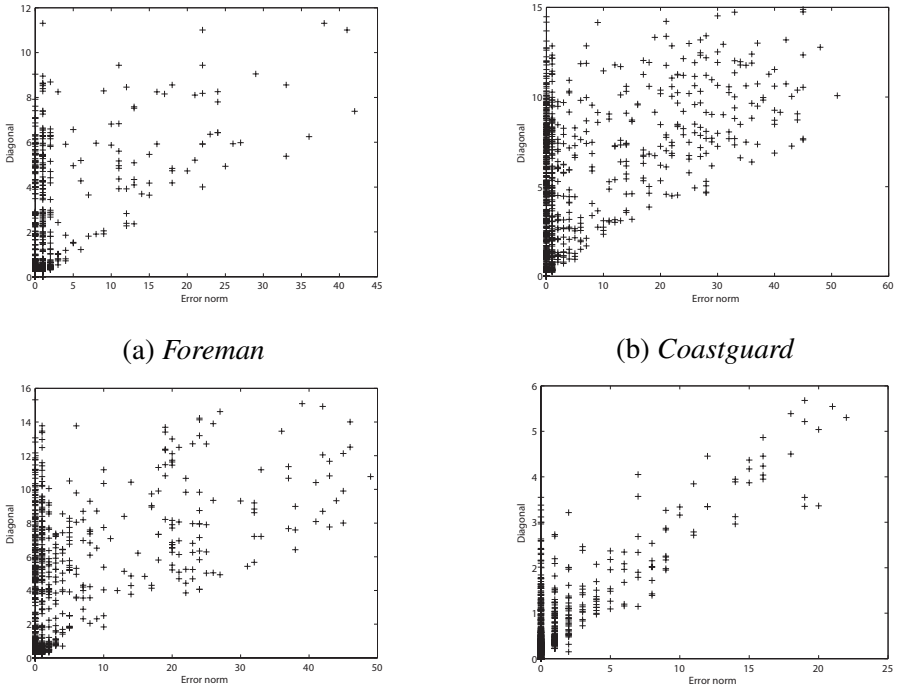
Figure 2.6: (a)∼(c) the observed MVs placed each at the top left corner of a block, (d)∼(f) SAD surfaces for the center block around the minimum point, (g)∼(i) 2D views of the SAD surfaces. The dots indicate the nine MVs. The red ones for those in $\{i\} \cup \{j | j \in \mathcal{N}_i, \epsilon_i(\mathbf{d}(j)) \leq 270\}$ and the blacks for the others.

that minimize the matching error best reflect the image characteristic of the ith block. Figure 2.6 shows examples. Figure 2.6-(a)$\sim$(c) show the observed MVFs, each vector placed at the top left corner of the 8$\times$8 block. Figure 2.6-(d)$\sim$(f) respectively show the SAD surfaces of the center blocks of Figure 2.6-(a)$\sim$(c) around the minimum points. Figure 2.6-(g)$\sim$(i) are the corresponding 2D views of Figure 2.6-(d)$\sim$(f), and the dots designate the 9 MVs in Figure 2.6-(a)$\sim$(c) on the SAD surfaces. The red dots indicate the MVs that result in the SADs smaller than a threshold $\lambda^{\mathbf{d}} = 270$ when applied to the center block. The threshold is determined by experiments. Note that Figure 2.6 is obtained from *Foreman* test video by full search block-matching with the block size 8$\times$8 pixels and the search range +/-24 in both directions.

Consider first Figure 2.6-(a) and (d) in which the center block contains complex textures. The corresponding SAD surface is likely to exhibit a prominent peak at the true motion so the MV pointing to the peak is usually considered reliable. The figures show that the observed MV at the minimum SAD point indeed coincides with the true. As indicated by a red dot in Figure 2.6-(g), the neighboring MVs, which usually minimize block-matching errors of the center block, unanimously point to the minimum.

The example of Figure 2.6-(b) and (e) has a SAD surface with a narrow valley in a certain direction which tends to occur in a block with a strong edge. The true motion is of course located in the valley. The observed MV of the block may point to the true or false motion and is considered reliable only in the direction perpendicular to that of the valley. Some vectors indicated by red dots point to the true motion, but others point to false motions in the valley, as are shown in Figure 2.6-(h).

The last example of Figure 2.6-(c), (f), and (i) shows a SAD surface with no noticeable peaks and with the observed MV of the center block pointing to a false motion. This occurs when the image has neither enough textures nor distinct features, and the MV of the block is regarded as unreliable. The vectors indicated by red dots point to



(a) *Foreman* (b) *Coastguard*

(c) *Mobile* (d) *City*

Figure 2.7: The observation error norm versus $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ which is calculated from the current block and its 8 neighbors.

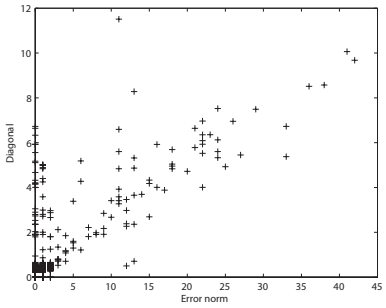
rather unpredictable places.

In short, the neighboring MVs that minimize block-matching errors of the center block reflect the prominence of the minimum peak of the SAD surface of the center block, which in turn reflects the local image characteristics.

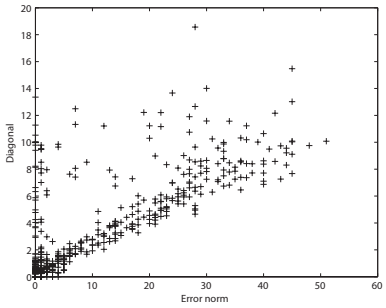
Based on such observations, the covariance matrix is estimated as follows: Let us assume that the ith block and its neighboring blocks in $\mathcal{N}_i^{\mathbf{d}} = \{l | l \in \mathcal{N}_i, \epsilon_i(\mathbf{d}(l)) \leq \lambda^{\mathbf{d}}\}$ are in the same moving object. As the true MVs belonging to the same moving object are the same, limited to translational motion only, the AGN model of (2.4) gives

$$\mathbf{d}(l) = \mathbf{u}(i) + \mathbf{e}(l), \;\; l \in \bar{\mathcal{N}}_i^{\mathbf{d}}, \tag{2.20}$$

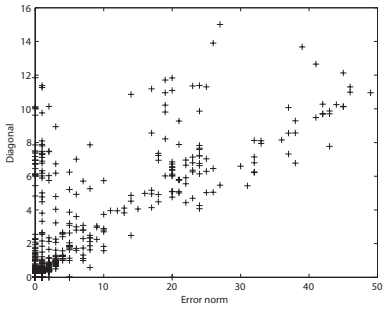
where $\bar{\mathcal{N}}_i^{\mathbf{d}} = \{\{i\} \cup \mathcal{N}_i^{\mathbf{d}}\}$. The errors $\mathbf{e}(l)$, $l \in \bar{\mathcal{N}}_i^{\mathbf{d}}$ are assumed to come from



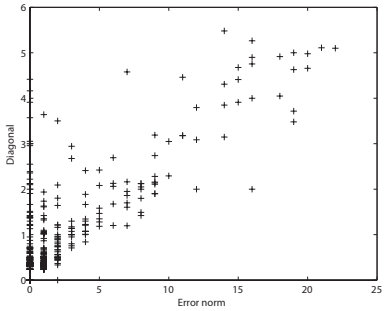
(a) *Foreman*



(b) *Coastguard*



(c) *Mobile*



(d) *City*

Figure 2.8: The observation error norm versus $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ which is calculated from the current block and $\mathcal{N}_i^{\mathbf{d}}$.

independent and identical random vectors based on our locally stationary AGN model. Then the maximum-likelihood estimate (MLE) of the covariance matrix is given by

$$\hat{C}_i^{\mathrm{ml}} = \frac{1}{L} \sum_{l \in \bar{\mathcal{N}}_i^{\mathbf{d}}} (\mathbf{d}(l) - \mathbf{m}_d)(\mathbf{d}(l) - \mathbf{m}_d)^T, \tag{2.21}$$

where $\mathbf{m}_d = \frac{1}{L} \sum_{l \in \bar{\mathcal{N}}_i^{\mathbf{d}}} \mathbf{d}(l)$ and $L = |\bar{\mathcal{N}}_i^{\mathbf{d}}|$. The unbiased estimate of C_i is then calculated as $\hat{C}_i = (L/(L-1)) \cdot \hat{C}_i^{\mathrm{ml}}$, which in turn takes part in the MAP estimate of (2.3).

The remaining problem is how to determine the threshold $\lambda^{\mathbf{d}}$. As illustrated in Figure 2.6, only the neighboring MVs that point to the minimum points of the SAD surface of the center block should be included in $\mathcal{N}_i^{\mathbf{d}}$ by $\lambda^{\mathbf{d}}$. Without knowing the SAD surface, however, it is not easy to determine a correct value of $\lambda^{\mathbf{d}}$, and computing the SAD surface for each block is very costly. So heuristic and trial-and-error approach is taken to find from experiments (see Figure 2.7 and Figure 2.8) that the following gives overall satisfactory results on our test videos:

$$\lambda^{\mathbf{d}} = \begin{cases} \max(\mathcal{S}_i^{\mathbf{d}}), & \text{if } \max(\mathcal{S}_i^{\mathbf{d}}) - \min(\mathcal{S}_i^{\mathbf{d}}) \leq 2|\mathcal{B}_i| \\ \min(\{\lambda_1, \lambda_2\}), & \text{otherwise,} \end{cases} \tag{2.22}$$

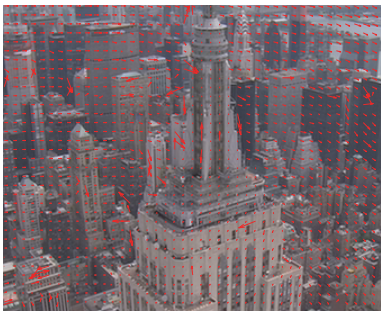
where $\lambda_1 = \text{median}(\mathcal{S}_i^{\mathbf{d}})$, $\lambda_2 = 0.5 \cdot (\max(\mathcal{S}_i^{\mathbf{d}}) + \min(\mathcal{S}_i^{\mathbf{d}}))$, and $\mathcal{S}_i^{\mathbf{d}} = \{\epsilon_i(\mathbf{d}(j)) \mid j \in \mathcal{N}_i\}$. The first condition of (2.22) corresponds to the case where all the MVs are used for the estimation as they give nearly the same SADs for the center block.

Note that considering only a portion of the neighboring MVs in the estimate (2.21) can be regarded as a robust parameter estimation [51], in which the observation that is close to the target model is given greater weight as compared to the one that is less so.

(a) *Foreman*

(b) *Coastguard*

(c) *Mobile*



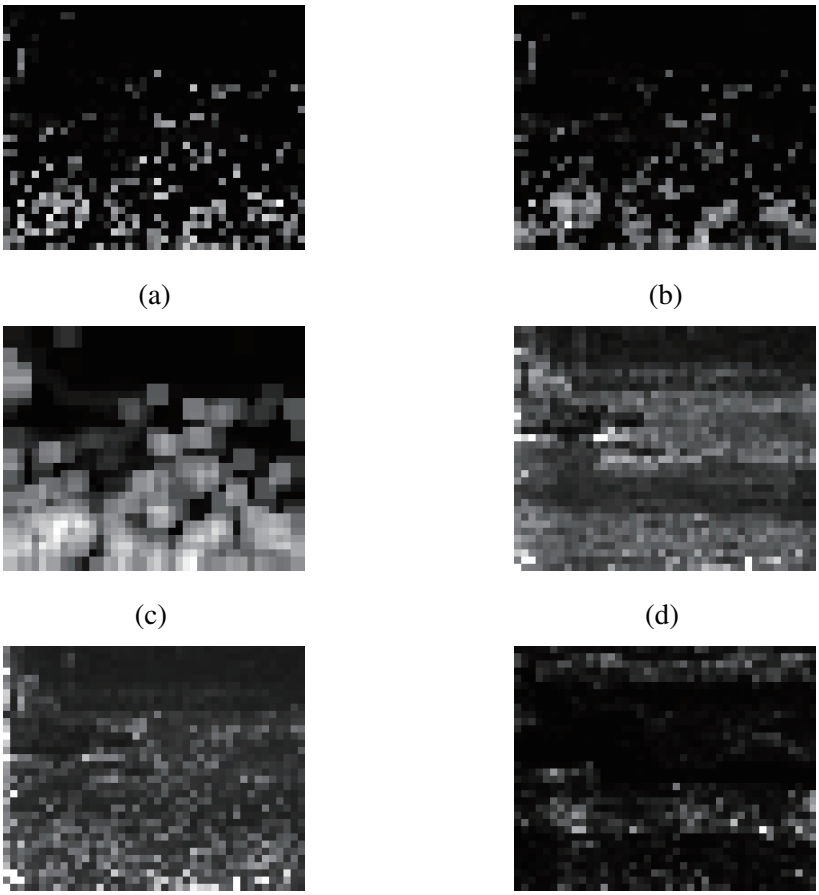
(d) *Mobile*

Figure 2.9: The observed MVF of (a) the 297-th frame of *Foreman*, (b) the 39-th frame of *Coastguard*, (c) the 11-th frame of *Mobile*, and (d) the 275-th frame of *City*.

2.2.2 Performance of the proposed reliability measure

In this section, it is shown that the estimated covariance matrix $\hat{C}_i$ is highly correlated with the reliability of the observation $\mathbf{d}(i)$ and the reliability measure $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ is compared with those of the existing measures. Figure 2.7 and 2.8 show how the diagonals of $\hat{C}_i$ are correlated with the ground-truth error of the observation. Note that Figure 2.7 and 2.8 are obtained from the test videos of Figure 2.9, which shows the observed backward MVFs, obtained by full search block-matching with the block size 8×8 pixels and the search range +/-24 in both directions. The ground-truth MVFs of the frames are obtained by manual selection with the help of VMF. The figure plots the l_1 norm of the observation error (i.e., $||\mathbf{d}(i) - \mathbf{u}(i)||_1$) of each block of the test videos versus $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$. Figure 2.8 is derived from the MVs of the current block and its neighbors in $\mathcal{N}_i^{\mathbf{d}}$ while Figure 2.7 use the whole neighborhood. It shows that the threshold $\lambda^{\mathbf{d}}$ of (2.22) properly discriminates neighbors to achieve stronger positive correlation.

The estimated reliability of the observed MVs in Figure 2.9-(b) is visualized in Figure 2.10. Figure 2.10-(a) presents in brightness the norm of the observation error vectors while (b)∼(f) show their estimated reliability. The brightness corresponds to unreliability. We find in the figures that our measure calculated from $\mathcal{N}_i^{\mathbf{d}}$ shows the strongest correlation with the norm of the observation error vector.



(a) (b) (c) (d) (e) (f)

Figure 2.10: Visualization in brightness of the unreliability of the observed MVF in Figure 2.9-(b). (a) norm of the observation error vectors, (b) $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ derived from $\mathcal{N}_i^{\mathbf{d}}$, (c) $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ derived from the whole neighbors, (d) minimum matching error, (e) block overlap [22], (f) principal curvature of the matching error surface [37].

2.3 Solution to MAP

Now the MAP estimate is obtained by combining (2.7) and (2.13) in (2.3), and consequently by minimizing the total energy given by

$$\begin{aligned}\mathcal{E}(\mathbf{u}) &= \sum_{i\in\mathcal{I}} V_T(\mathbf{u}(i), \mathcal{L}_i) + K \\ &= \sum_{i\in\mathcal{I}} \Big[\frac{1}{2}[\mathbf{d}(i)-\mathbf{u}(i)]^T C_i^{-1}[\mathbf{d}(i)-\mathbf{u}(i)] + \gamma \sum_{j\in\mathcal{N}_i} \mu_{ij}||\mathbf{u}(i)-\mathbf{u}(j)||^2 \\ &+ \frac{1}{2}\ln \eta_i\kappa_i\Big] + K, \end{aligned} \tag{2.23}$$

where $\mathcal{L}_i = \{\mathbf{u}(j)|j \in \mathcal{N}_i\}$ and $K = \ln Z + |\mathcal{I}|\ln 2\pi$ is a constant. The iterative conditional mode (ICM) algorithm [47] is used to minimize the total energy $\mathcal{E}(\mathbf{u})$. Given $\mathbf{d}(i)$ and $\mathcal{L}_i^k = \{\mathbf{u}^k(j)|j \in \mathcal{N}_i\}$, where $\mathbf{u}^k(j) = [u_x^k(j)\ u_y^k(j)]^T$ denotes the kth estimate of $\mathbf{u}(j)$, the algorithm sequentially updates $\mathbf{u}^k(i)$ into $\mathbf{u}^{k+1}(i)$ by minimizing for each i the overall potential function $V_T(\mathbf{u}(i), \mathcal{L}_i^k)$, that is,

$$\mathbf{u}^{k+1}(i) = \underset{\mathbf{u}(i)\in\Theta}{\operatorname{argmin}}\, V_T(\mathbf{u}(i), \mathcal{L}_i^k),\ \forall i \in \mathcal{I}, \tag{2.24}$$

where Θ is the set of all candidate MVs of the ith block. Since it is a convex function of $\mathbf{u}(i)$, the vector that minimizes it can be found by solving $\frac{\partial V_T(\mathbf{u}(i),\mathcal{L}_i^k)}{\partial \mathbf{u}(i)} = 0$, which gives rise to the updating equations:

$$u_x^{k+1}(i) = \frac{d_x(i) + 2\gamma\eta_i \sum_{j\in\mathcal{N}_i} \mu_{ij} u_x^k(j)}{1 + 2\gamma\eta_i}, \tag{2.25a}$$

$$u_y^{k+1}(i) = \frac{d_y(i) + 2\gamma\kappa_i \sum_{j\in\mathcal{N}_i} \mu_{ij} u_y^k(j)}{1 + 2\gamma\kappa_i}. \tag{2.25b}$$

Note that μ_{ij} is regarded as a constant while solving $\frac{\partial V_T(\mathbf{u}(i),\mathcal{L}_i^k)}{\partial \mathbf{u}(i)} = 0$ since it is not a function of $\mathbf{u}(i)$.

The MAP estimate of the true MVF is obtained by iterating (2.25) with the initial estimate $\mathbf{u}^0 = \mathbf{d}$ and with μ_{ij}'s updated using $\mathbf{u}^k$ until $|\mathcal{E}(\mathbf{u}^{k+1}) - \mathcal{E}(\mathbf{u}^k)|$ drops below a pre-defined threshold. Also, $\eta_i = \kappa_i = \sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ is used in (2.25) for the

experiments in section 2.4.

2.4 Relation to Previous Works

In this section, the relationship between the proposed and the existing MVF regularization algorithms is discussed. Anandan first introduced a regularization framework into MVF smoothing [37]. The cost function to be minimized with respect to $\mathbf{u}$ for the estimation has the form

$$C(\mathbf{u}) = \lambda\Delta(\mathbf{u}, \mathbf{d}) + \Phi(\mathbf{u}) \tag{2.26}$$

where $\Delta(\mathbf{u}, \mathbf{d})$ is a measure of the difference between $\mathbf{u}$ and the observation $\mathbf{d}$, $\Phi(\mathbf{u})$ is a regularization functional quantifying the smoothness of the MVF, and λ is a constant. Different algorithms are formulated by different choices of $\Delta(\mathbf{u}, \mathbf{d})$ and $\Phi(\mathbf{u})$ [39]. The following iterative update equation for the kth estimate $\mathbf{u}^k$ of $\mathbf{u}$ can be obtained from the cost suggested in [37]:

$$\begin{aligned}\mathbf{u}^{k+1}(i) = \bar{\mathbf{u}}^k(i) &+ c_{\max}((\mathbf{d}(i) - \bar{\mathbf{u}}^k(i))^T\mathbf{e}_{\max})\mathbf{e}_{\max} \\ &+ c_{\min}((\mathbf{d}(i) - \bar{\mathbf{u}}^k(i))^T\mathbf{e}_{\min})\mathbf{e}_{\min}, \quad \forall i \in \mathcal{I},\end{aligned} \tag{2.27}$$

where $c_{\max}$ and $c_{\min}$ are the weights related to the principal curvatures of the matching error surface of the ith block at the minimum point, $\mathbf{e}_{\max}$ and $\mathbf{e}_{\min}$ are the 2-D vectors corresponding to the principal curvature axes, and $\bar{\mathbf{u}}^k(i) = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\mathbf{u}^k(j)$. For better understanding, let's consider a special case when $\mathbf{e}_{\max} = [1\ 0]^T$ and $\mathbf{e}_{\min} = [0\ 1]^T$. Then, (2.27) becomes

$$u_x^{k+1}(i) = c_{\max}d_x(i) + (1 - c_{\max})\bar{u}_x^k(i) \tag{2.28a}$$

$$u_y^{k+1}(i) = c_{\min}d_y(i) + (1 - c_{\min})\bar{u}_y^k(i). \tag{2.28b}$$

We can see that the $(k+1)$th estimate of the true MV is a weighted sum of the observation and the neighborhood average of the kth estimate. In this formulation, Zheng [39]

incorporated the minimum matching errors of the neighboring blocks in the calculation of $\bar{\mathbf{u}}^k(i)$. Sohn et al. [40] also proposed a similar MVF regularization algorithm dealing especially with problems arising in areas with spatially periodic patterns.

The iterative update equations (2.25) resembles (2.28) in that the $(k+1)$th estimate of the true MV is a weighted sum of the observation and the average of the neighbors in the kth estimate. The differences however are in the weights and the way to compute the neighborhood average. In trading off the observation against the neighborhood average, the proposed algorithm uses the variance of the observations $\{\mathbf{d}(l)|l \in \bar{\mathcal{N}}_i^{\mathbf{d}}\}$, which reflects the prominence of the minimum peak of the matching error surface. Sohn et al. on the other hand compute the weights using not only the block-matching errors of the current block but also the sum of distances between $\bar{\mathbf{u}}^k(i)$ and the vectors in $\{\mathbf{d}(l)|l \in \mathcal{N}_i\}$, in order to deal with spatially periodic patterns. Anandan and Zheng's utilization of the principal curvatures of the matching error surface at the minimum point is derived from the observation that the sharpness of the matching error surface at the minimum point generally correlates with the reliability of the MV pointing to that point. But the sharpness measure can produce undesirable results when the matching error surface has multiple prominent peaks, for example, in an area with periodic patterns. Multiple prominent peaks also invalidate the minimum matching error value itself as a reliability measure.

For the calculation of the average of the neighbors, the proposed algorithm uses $\bar{\mathbf{u}}^k(i) = \sum_{j \in \mathcal{N}_i} \mu_{ij} \mathbf{u}^k(j)$ where μ_{ij} is the weight defined in (2.15). The weight measures how closely the neighboring MV $\mathbf{u}^k(j)$ corresponds to the true motion of the ith block. In contrast, Zheng let the weight be negatively correlated with the minimum matching error of the jth block, thereby weakening the influence of false motions in the occluded area where the matching errors are commonly large. Anandan and Sohn's isotropic neighborhood average can be expressed as $\bar{\mathbf{u}}^k(i) = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \mathbf{u}^k(j)$, where all the neighbors contribute equally.

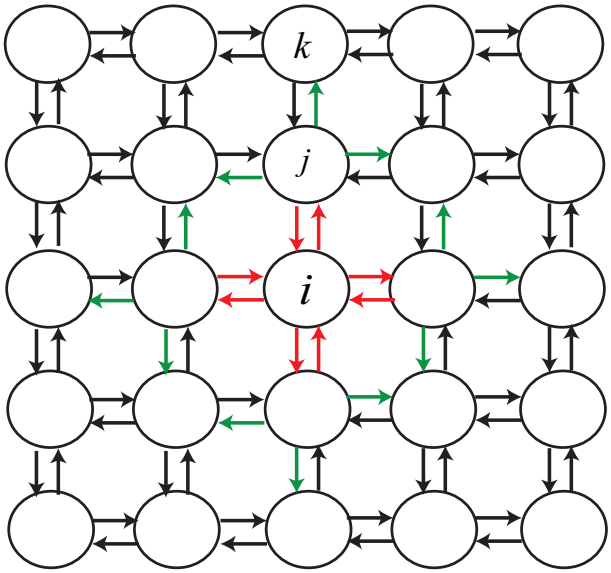


Figure 2.11: Interaction of random vectors in a vector field with the distribution (2.13).

2.5 Properties of Proposed Prior Distribution

This section discusses the properties of the prior distribution of (2.13). Let $\mathbf{U} = \{\mathbf{U}_i | i \in \mathcal{S}\}$ be a family of 2-D random vectors defined on the set of sites $\mathcal{S} = \{i | i = 1, ..., N\}$ in which each random vector $\mathbf{U}_i$ takes a value $\mathbf{u}_i$ in an alphabet $\mathcal{L}$. Then a random vector field with the distribution of (2.13) shows the following two properties.

a) positivity : Since the distribution of (2.13) is an exponential function, $p(\mathbf{U} = \mathbf{u}|I_p, I_c) \triangleq p(\mathbf{u}|I_p, I_c) > 0, \forall \mathbf{u} \in \mathbb{U}$ where $\mathbb{U}$ is the set of all possible configurations of $\mathbf{U}$.

b) semi-markovianity : Consider the conditional probability

$$p(\mathbf{u}_i|\mathbf{u}_{\mathcal{S}-\{i\}}) = \frac{p(\mathbf{u}_i, \mathbf{u}_{\mathcal{S}-\{i\}})}{p(\mathbf{u}_{\mathcal{S}-\{i\}})} = \frac{p(\mathbf{u})}{\sum_{\mathbf{u}'_i \in \mathcal{L}} p(\mathbf{u}')}, \tag{2.29}$$

where $\mathcal{S} - \{i\}$ is the set difference, $\mathbf{u}_{\mathcal{S}-\{i\}}$ denotes the set of labels at the sites in $\mathcal{S} - \{i\}$, and $\mathbf{u}'$ is any configuration that agrees with $\mathbf{u}$ at all sites except possibly i.

Let us express (2.13) as

$$p(\mathbf{u}|I_p, I_c) \triangleq p(\mathbf{u}) = \frac{1}{Z} \exp\Big(- \gamma \sum_{c \in \mathcal{C}} V_c(\mathbf{u}) \Big), \tag{2.30}$$

where c is a clique, $\mathcal{C}$ is the set of all possible cliques, and $V_c(\mathbf{u}) = \mu_{ij} ||\mathbf{u}_i - \mathbf{u}_j||^2$ is the potential function for the clique $c = \{i, j\}$. Then $p(\mathbf{u}_i|\mathbf{u}_{\mathcal{S}-\{i\}})$ can be expressed as

$$p(\mathbf{u}_i|\mathbf{u}_{\mathcal{S}-\{i\}}) = \frac{\exp\Big(- \sum_{c \in \mathcal{C}} V_c(\mathbf{u}) \Big)}{\sum_{\mathbf{u}'_i \in \mathcal{L}} \exp\Big(- \sum_{c \in \mathcal{C}} V_c(\mathbf{u}') \Big)}. \tag{2.31}$$

The set $\mathcal{C}$ of cliques divides into $\mathcal{A}$ and $\mathcal{B}$, with $\mathcal{A}$ consisting of cliques containing i and $\mathcal{B}$ not containing i. $\mathcal{B}$ again divides into $\mathcal{B}_i$ of the cliques whose potential (owing to the weights μ_{jk}'s) is affected by $\mathbf{u}_i$ and $\bar{\mathcal{B}}_i = \mathcal{B} - \mathcal{B}_i$. Figure 2.11 shows an example of the sets when the 4 neighborhood system is considered, where a circle denotes a site and the arrow from *i* to *j* indicates the clique $c = \{i, j\}$. The sets $\mathcal{A}$, $\mathcal{B}_i$, and $\bar{\mathcal{B}}_i$ respectively comprise the red, green, and black arrows. Then (2.31) can be written as

$$p(\mathbf{u}_i|\mathbf{u}_{\mathcal{S}-\{i\}}) = \frac{\exp\Big(- \sum_{c \in \mathcal{A}} V_c(\mathbf{u}) - \sum_{c \in \mathcal{B}_i} V_c(\mathbf{u}) - \sum_{c \in \bar{\mathcal{B}}_i} V_c(\mathbf{u}) \Big)}{\sum_{\mathbf{u}'_i \in \mathcal{L}} \exp\Big(- \sum_{c \in \mathcal{A}} V_c(\mathbf{u}') - \sum_{c \in \mathcal{B}_i} V_c(\mathbf{u}') - \sum_{c \in \bar{\mathcal{B}}_i} V_c(\mathbf{u}') \Big)} \tag{2.32}$$

Since $V_c(\mathbf{u}) = V_c(\mathbf{u}')$ for any clique c that does not contain i or whose potential is not affected by i, $\exp(-\sum_{\bar{\mathcal{B}}_i} V_c(\mathbf{u}))$ cancels from both the numerator and denominator to result in

$$p(\mathbf{u}_i|\mathbf{u}_{\mathcal{S}-\{i\}}) = \frac{\exp\Big(- \sum_{c \in \mathcal{A}} V_c(\mathbf{u}) - \sum_{c \in \mathcal{B}_i} V_c(\mathbf{u}) \Big)}{\sum_{\mathbf{u}'_i \in \mathcal{L}} \exp\Big(- \sum_{c \in \mathcal{A}} V_c(\mathbf{u}') - \sum_{c \in \mathcal{B}_i} V_c(\mathbf{u}') \Big)}. \tag{2.33}$$

The sites involved in the cliques of $\mathcal{B}_i$ are the neighbors to i as seen in the figure, so the probability of a random vector depends not only on its neighbors but also on the neighbors of the neighbors. The proposed distribution is therefore different from Gibbs distribution in which the probability of a random vector depends only on the neighbors.

This derivation is based on the proof of Markov-Gibbs equivalence detailed in [53].

Chapter 3

Multiple Hypotheses Bayesian Approach

In the previous chapter, the true motion vector field (MVF) of a video frame is estimated from its observed MVF under the assumption that the observed MVF, which is the result of a block-based motion estimation (BME), is a degraded version of the true MVF by locally stationary additive Gaussian noise (AGN). In this chapter, multiple observations, which are the results of a BME incorporating multiple block sizes for block-matching, are considered for the estimation of the true MVF.

3.1 Problem Formulation

Given two adjacent frames I_p (the previous frame) and I_c (the current frame) of a video, let $\mathbf{d}_k = \{\mathbf{d}_k(i)|i \in \mathcal{I}\}$ denotes the 2-D backward MVF obtained by the kth BME algorithm where $k = 1, ..., M$, $\mathbf{d}_k(i)$ ($\in \mathbb{R}^2$) denotes the kth motion vector (MV) of the ith block of the frame I_c pointing to a matching block in I_p, and $\mathcal{I}$ and $M \geq 1$ respectively are the set of all block indices in a frame and the number of the MVFs. Let's assume that $\mathbf{d}_k$, called the kth *observed MVF*, or simply the kth *observation*, is a degraded version of the true MVF $\mathbf{u} = \{\mathbf{u}(i)|i \in \mathcal{I}\}$, where $\mathbf{u}(i) = [u_x(i)\ u_y(i)]^T$ ($\in \mathbb{R}^2$) is the true MV of the ith block. The degradation process is expressed as

$$\mathbf{d}_k = \varphi_k(\mathbf{u}),\ k = 1, ..., M \tag{3.1}$$

for some kth degradation operation $\varphi_k(\cdot)$. Under this assumption, the true MVF $\mathbf{u}$ is estimated from the observations $\mathbf{d}_k,\ k = 1, ..., M$ by maximizing the posterior probability

$$p(\mathbf{u} \mid \mathbf{d}_1, ..., \mathbf{d}_M, I_p, I_c) = \frac{p(\mathbf{d}_1, ..., \mathbf{d}_M \mid \mathbf{u}, I_p, I_c)p(\mathbf{u} \mid I_p, I_c)p(I_p, I_c)}{p(\mathbf{d}_1, ..., \mathbf{d}_M, I_p, I_c)}. \tag{3.2}$$

Therefore, the MAP estimate $\hat{\mathbf{u}}$ of $\mathbf{u}$ can be obtained by

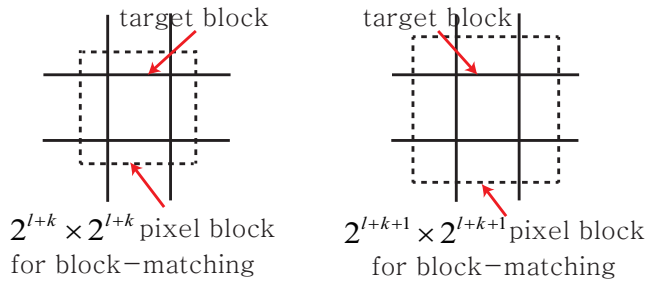
$$\hat{\mathbf{u}} = \underset{\mathbf{u} \in \mathcal{U}}{\operatorname{argmax}} [p(\mathbf{d}_1, ..., \mathbf{d}_M \mid \mathbf{u}, I_p, I_c)p(\mathbf{u} \mid I_p, I_c)], \tag{3.3}$$

where $\mathcal{U}$ denotes the set of all candidate MVFs for I_c. Note that $p(\mathbf{d}_1, ..., \mathbf{d}_M, I_p, I_c)$ and $p(I_p, I_c)$ are ignored because they are constants with respect to the unknown.

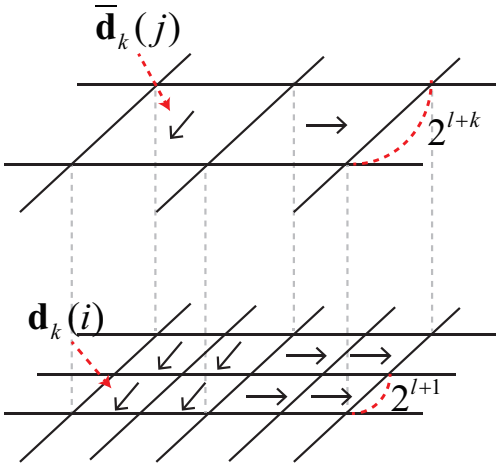
3.1.1 Proposed observation likelihood

In this dissertation, it is assumed that the multiple observed MVFs are obtained by a BME algorithm incorporating multiple block sizes for block-matching. Let N_1 denote the dimension of the true MVF, which is the same as the number of non-overlapping $2^{l+1} \times 2^{l+1}$ pixel blocks in a frame, where $l \geq 1$ is an integer. Then, the kth observation of the true MV of a $2^{l+1} \times 2^{l+1}$ pixel block is obtained by full search block-matching with the matching block of size $2^{l+k} \times 2^{l+k}$ pixels where $k \geq 1$. Figure 3.1-(a) shows an example of obtaining multiple observed MVs using different block sizes, called *overlapping matching* method. This approach, however, requires high computational load since, for each target block of size $2^{l+1} \times 2^{l+1}$ pixels, M times of full search block-matching should be done, where M denotes the number of the observed MVFs.

In order to reduce the computational load, obtaining multiple observations using *non-overlapping matching* method is suggested as seen in Figure 3.1-(b). The kth observed MVF $\mathbf{d}_k$ of the true MVF $\mathbf{u}$ is obtained as follow. Let $\bar{\mathbf{d}}_k$ denote the parent



(a) Overlapping matching

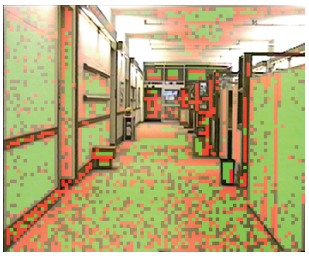


(b) Non-overlapping matching

Figure 3.1: Examples of obtaining multiple observed MVFs.

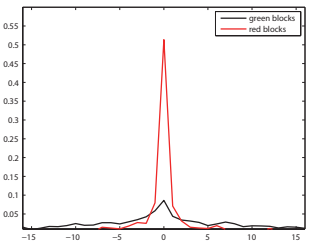
MVF of dimension N_k, which is the number of non-overlapping $2^{l+k} \times 2^{l+k}$ pixel blocks in a frame. $\bar{\mathbf{d}}_k$ is obtained by full search block-matching with non-overlapping $2^{l+k} \times 2^{l+k}$ pixel matching block. Note that $N_1 \geq N_k$ holds since $k \geq 1$. Then the ith MV $\mathbf{d}_k(i)$ of $\mathbf{d}_k$ is obtained by letting $\mathbf{d}_k(i) = \bar{\mathbf{d}}_k(j)$ where the ith $2^{l+1} \times 2^{l+1}$ pixel block is a sub-block of the jth $2^{l+k} \times 2^{l+k}$ pixel block.

Now let us look into the statistical nature of observation errors to define the degradation process in (3.1). In Figure 3.2-(a) and (b), blocks of a similar texture are presented in the same color. The green and red correspond to the low and mid texture content, respectively. The amount of texture is measured by the variance of the pix-

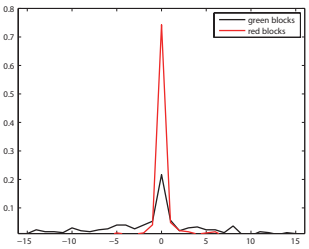


(a) 4×4 block

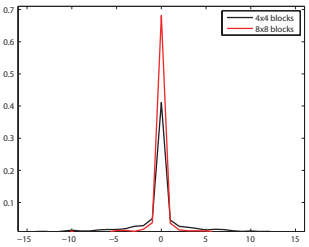
(b) 8×8 block



(c) 4×4 block



(d) 8×8 block



(e)

Figure 3.2: The normalized histograms of x-axis components of the observation error vectors. (a), (b) texture contents classification; green, red correspond to low and mid texture, respectively, (c), (d) the normalized histograms of the error vectors color-coded for the respective regions, (e) the normalized histograms of the error vectors for the full frame.

els in the block. Note that 4×4 and 8×8 pixel blocks are used for obtaining Figure 3.2-(a) and (b), respectively. Figure 3.2-(c) and (d) respectively show the normalized histograms of x-axis components of the observation error vectors color-coded for the respective regions of Figure 3.2-(a) and (b). The observations are obtained by full search block-matching with block size 4×4 pixels (for Figure 3.2-(c)) and 8×8 pixels (for Figure 3.2-(d)) and search range +/-24 both in vertical and horizontal direction at integer pixel accuracy. As we observed in section 2.1.1, in both cases, the texture content of a block is inversely correlated with the magnitude of the error vector. In addition, blocks of the same color are clustered together. Figure 3.2-(e) shows the normalized histograms of the error vectors for the full frame. We can see in the figure that the width of the histogram originating from 8×8 block is more narrow than that of the histogram originating from 4×4 block. This is because a larger block usually includes more texture contents for the matching than a smaller block.

Based on such observations, the degradation process $\varphi_k(\cdot)$ in (3.1) is modeled as locally stationary additive Gaussian noise (AGN) such that

$$\mathbf{d}_k(i) = \mathbf{u}(i) + \mathbf{e}_k(i), \;\; i \in \mathcal{I}, \; k = 1, ..., M, \tag{3.4}$$

where $\mathbf{e}_k(i)$ ($\in \mathbb{R}^2$) is a realization of a 2-D random noise vector following $\mathcal{N}(\mathbf{0}, c_{ki}^{-1} \cdot \mathbf{I})$. $\mathbf{0}$ and $\mathbf{I}$ are the 2×1 zero vector and the 2×2 identity matrix, respectively, and c_{ki}^{-1} is the noise variance of the kth observation of the ith block. As mentioned in the previous chapter, locally stationary AGN means that c_{ki}^{-1}, $i \in \mathcal{I}$ may not show the same value over the whole frame but may show almost the same value in local areas of the frame. In addition, we can expect from Figure 3.2-(e) that $c_{ki}^{-1} \geq c_{mi}^{-1}$ usually holds when $m \geq k$.

It is first assumed that $\mathbf{d}_k$, $k = 1, ..., M$ are conditionally independent given $\mathbf{u}$, I_p, and I_c so the joint probability $p(\mathbf{d}_1, ..., \mathbf{d}_M \mid \mathbf{u}, I_p, I_c)$ can be expressed as

$$p(\mathbf{d}_1, ..., \mathbf{d}_M \mid \mathbf{u}, I_p, I_c) = \prod_{k=1}^{M} p(\mathbf{d}_k \mid \mathbf{u}, I_p, I_c). \tag{3.5}$$

Note that $\mathbf{d}_k$ and $\mathbf{d}_m$ ($m > k$), in reality, are not independent each other since the same region of a video frame is utilized for obtaining $\mathbf{d}_k(i)$ and $\mathbf{d}_m(i)$ ($i \in \mathcal{I}$) through block-matching. However the conditional independence is assumed in this dissertation for the sake of simple analysis. Next two assumptions are further made in calculating $p(\mathbf{d}_k \mid \mathbf{u}, I_p, I_c)$. Assumed are that the observations $\mathbf{d}_k(i)$, $i \in \mathcal{I}$, are conditionally independent given $\mathbf{u}$, I_p, and I_c, and that each $\mathbf{d}_k(i)$ is independent of $\mathbf{u}(j)$, $j \neq i$, $j \in \mathcal{I}$ so that $p(\mathbf{d}_k(i)|\mathbf{u}, I_p, I_c) = p(\mathbf{d}_k(i)|\mathbf{u}(i), I_p, I_c)$. The above two assumptions do not always hold in reality except for $\mathbf{d}_1$. This is because the same region of a video frame is utilized for obtaining neighboring MVs $\mathbf{d}_k(i)$ and $\mathbf{d}_m(j)$ ($i, j \in \mathcal{I}$) through block-matching. For the sake of simple analysis, however, the conditional independence is assumed in this dissertation. It then follows that $p(\mathbf{d}_k \mid \mathbf{u}, I_p, I_c) = \prod_{i=1}^{N} p(\mathbf{d}_k(i) \mid \mathbf{u}(i), I_p, I_c)$, and our observation model (3.4) finally lead us to the following observation likelihood

$$
\begin{aligned}
p(\mathbf{d}_1, ..., \mathbf{d}_M \mid \mathbf{u}, I_p, I_c) &= \prod_{k=1}^{M} \prod_{i \in \mathcal{I}} p(\mathbf{d}_k(i) \mid \mathbf{u}(i), I_p, I_c) \\
&= \Big(\prod_{k=1}^{M} \prod_{i \in \mathcal{I}} \frac{c_{ki}}{2\pi} \Big) \exp\Big(- \sum_{k=1}^{M} \sum_{i \in \mathcal{I}} \frac{c_{ki} \|\mathbf{d}_k(i) - \mathbf{u}(i)\|^2}{2} \Big).
\end{aligned} \tag{3.6}
$$

Note that the noise variance c_{ki}^{-1} corresponds to our confidence in $\mathbf{d}_k(i)$ and is estimated from the kth observation $\mathbf{d}_k$ as will be described in section 3.3.

3.1.2 Prior distribution of true motion vector field

In this chapter, the true MVF is assumed to follow the prior distribution proposed in chapter 2.

3.2 MAP Solution

Now the MAP estimate is obtained by combining (3.6) and (2.13) in (3.3), and consequently by minimizing of the total energy given by

$$\mathcal{E}(\mathbf{u}) = \sum_{i \in \mathcal{I}} V_T(\mathbf{u}(i), \mathcal{L}_i) + R \tag{3.7}$$

where

$$V_T(\mathbf{u}(i), \mathcal{L}_i) = \sum_{k=1}^{M} \frac{c_{ki} ||\mathbf{d}_k(i) - \mathbf{u}(i)||^2}{2} + \gamma \sum_{j \in \mathcal{N}_i} \mu_{ij} ||\mathbf{u}(i) - \mathbf{u}(j)||^2, \tag{3.8}$$

$\mathcal{L}_i = \{\mathbf{u}(j) | j \in \mathcal{N}_i\}$, and $R = \frac{1}{2} \sum_k \sum_i \ln \frac{c_{ki}}{2\pi} + \ln Z$ is a constant. We use the iterative conditional mode algorithm [47] to minimize the total energy $\mathcal{E}(\mathbf{u})$. The algorithm sequentially updates the tth estimate $\mathbf{u}^t(i)$ into the (t+1)th estimate $\mathbf{u}^{t+1}(i)$ by minimizing for each i the overall potential function $V_T(\mathbf{u}(i), \mathcal{L}_i^t)$, that is,

$$\mathbf{u}^{t+1}(i) = \underset{\mathbf{u}(i) \in \Theta}{\operatorname{argmin}}\, V_T(\mathbf{u}(i), \mathcal{L}_i^t),\ i \in \mathcal{I}, \tag{3.9}$$

where Θ is the set of all candidate MVs for the ith block, $\mathcal{L}_i^t = \{\mathbf{u}^t(j) | j \in \mathcal{N}_i\}$. Since $V_T(\mathbf{u}(i), \mathcal{L}_i^t)$ is a convex function of $\mathbf{u}(i)$, the vector that minimizes it can be found by solving $\frac{\partial V_T(\mathbf{u}(i), \mathcal{L}_i^t)}{\partial \mathbf{u}(i)} = 0$, which gives rise to the updating equation:

$$\mathbf{u}^{t+1}(i) = \frac{\sum_{k=1}^{M} c_{ki} \mathbf{d}_k(i) + 2\gamma \sum_{j \in \mathcal{N}_i} \mu_{ij} \mathbf{u}^t(j)}{2\gamma + \sum_{k=1}^{M} c_{ki}}. \tag{3.10}$$

It is seen in (3.10) that the $(t+1)$th estimate $\mathbf{u}^{t+1}(i)$ of the true MV of the ith block is obtained by combining its multiple observed MVs $\mathbf{d}_k(i)$, $k = 1, ..., M$ and previously estimated neighboring MVs $\{\mathbf{u}_j^t | j \in \mathcal{N}_i\}$ according to the noise variances c_{ki}^{-1}, $k = 1, ..., M$ for the observed MVs.

The HME algorithm proposed by Choi et al.[28] estimates the true MV of a block by choosing the most reliable MV among multiple observed MVs, which is obtained by full search block-matching incorporating multiple block sizes for block-matching.

This approach has a drawback that it produces a false estimate of the true MV when all of the observed MVs are unreliable. In contrast, the proposed algorithm utilizes the previously estimated neighboring MVs for the current estimation when the observed MVs seem unreliable. That is, $\mathbf{u}^{t+1}(i) \approx \sum_{j\in\mathcal{N}_i} \mu_{ij}\mathbf{u}^t(j)$ when the noise variances are sufficiently large so $\sum_{k=1}^{M} c_{ki} \ll 2\gamma$ is satisfied.

Finally note that, in this chapter, the true MVF is estimated by repeating (3.10) for all $i \in \mathcal{I}$ with an initial estimate $\mathbf{u}^0 = \mathbf{d}_1$ where $\mathbf{d}_1$ denotes the observed MVF obtained using the smallest block.

3.3 Adaptive Adjustment of Estimated Noise Variances

The noise variance c_{ki}^{-1}, $i \in \mathcal{I}$, $k = 1, ..., M$, in improving the estimate of the true MV of the ith block, trades off the contribution of the observation $\mathbf{d}_k(i)$ against that of the neighboring MVs as seen in (3.10). The trade-off should naturally depend on the confidence we have on the observation. In the previous chapter, we estimate the noise variances directly from the observation through maximum-likelihood estimation. In this chapter, the noise variances are estimated in the same manner. The following describes the estimation of the noise variance c_{ki}^{-1} for the kth observation $\mathbf{d}_k(i)$ of the true MV $\mathbf{u}(i)$.

Let us first assume that the ith $2^{l+1} \times 2^{l+1}$ pixel block of a frame is a sub-block of the jth $2^{l+k} \times 2^{l+k}$ pixel block of the frame. Also let $\bar{\mathbf{d}}_k$ be the parent MVF of $\mathbf{d}_k$ where $\bar{\mathbf{d}}_k$ is obtained by full search block-matching with non-overlapping $2^{l+k} \times 2^{l+k}$ pixel matching block. Then the estimate $\hat{c}_{ki}^{-1}$ of the noise variance c_{ki}^{-1} is obtained by letting $\hat{c}_{ki}^{-1} = \hat{\bar{c}}_{kj}^{-1}$. Here the estimate $\hat{\bar{c}}_{kj}^{-1}$ of the noise variance $\bar{c}_{kj}^{-1}$ for $\bar{\mathbf{d}}_k(j)$ is obtained by $\hat{\bar{c}}_{kj}^{-1} = \frac{1}{2} trace\{C_{kj}\}$ where

$$C_{kj} = \frac{1}{L-1} \sum_{l\in\bar{\mathcal{N}}_j^{\mathbf{d}}} (\bar{\mathbf{d}}_k(l) - \mathbf{m}_d)(\bar{\mathbf{d}}_k(l) - \mathbf{m}_d)^T, \tag{3.11}$$

$\mathbf{m}_d = \frac{1}{L}\sum_{l\in\bar{\mathcal{N}}_j^{\mathbf{d}}} \bar{\mathbf{d}}_k(l)$, $L = |\bar{\mathcal{N}}_j^{\mathbf{d}}|$, $\mathcal{N}_j^{\mathbf{d}} = \{l | l \in \mathcal{N}_j, \epsilon_j(\bar{\mathbf{d}}_k(l)) \le \lambda^{\mathbf{d}}\}$, and $\bar{\mathcal{N}}_j^{\mathbf{d}} =$

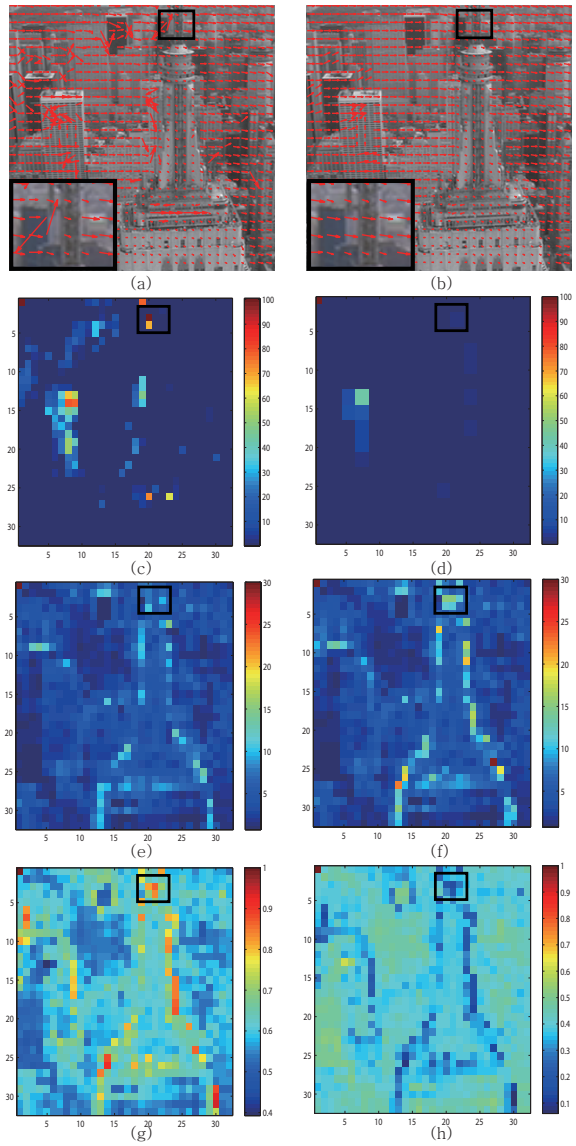


Figure 3.3: (a) the observed backward MVF $\mathbf{d}_1$ obtained by 8×8 pixel matching block, (b) the observed backward MVF $\mathbf{d}_2$ obtained by 16×16 pixel matching block, (c) the estimated noise variances $\hat{c}_{1i}^{-1}$'s of $\mathbf{d}_1$, (d) the estimated noise variances $\hat{c}_{2i}^{-1}$'s of $\mathbf{d}_2$, (e) MAD values of 8×8 blocks obtained using $\mathbf{d}_1$, (f) MAD values of 8×8 blocks obtained using $\mathbf{d}_2$, (g) ω_{1i}'s calculated with $\rho_1 = 1.0$, (h) ω_{2i}'s calculated with $\rho_2 = 1.3$.

$\{\{j\} \cup \mathcal{N}_j^{\mathbf{d}}\}$. Note that $\epsilon_j(\bar{\mathbf{d}}_k(l))$ denotes the SAD of the jth $2^{l+k} \times 2^{l+k}$ pixel block when $\bar{\mathbf{d}}_k(l)$ is applied. $\lambda^{\mathbf{d}}$ for $\mathcal{N}_j^{\mathbf{d}}$ is obtained from $\{\bar{\mathbf{d}}_k(l)|l \in \mathcal{N}_j\}$ using (2.22) with $|\mathcal{B}_j| = 2^{l+k} \times 2^{l+k}$.

Figure 3.3-(a) and (b), respectively, show the observed backward MVFs $\mathbf{d}_1$ and $\mathbf{d}_2$ that are obtained by a full search block-matching with the matching blocks of size 8×8, 16×16 pixels, and Figure 3.3-(c) and (d) present the noise variances estimated from $\mathbf{d}_1$ and $\mathbf{d}_2$ using (3.11), respectively. It is shown in the figures that the noise variance $\hat{c}_{2i}^{-1}$ for $\mathbf{d}_2$ is normally smaller than the variance $\hat{c}_{1i}^{-1}$ for $\mathbf{d}_1$. This is the consequence that the 16×16 block used for obtaining $\mathbf{d}_2$ includes more texture contents crucial for correct matching than 8×8 block for $\mathbf{d}_1$ so reliable MVs is obtained through block-matching. Thus if we simply let $c_{ki} = \hat{c}_{ki}$ in (3.10), the estimate $\mathbf{u}^t(i)$ of the true MV $\mathbf{u}(i)$ will get influenced mainly by $\mathbf{d}_2(i)$. This natural approach, however, can produce undesirable estimation results around motion boundaries. We can use Figure 3.3 to explain the problem and its solution.

Consider that the ith block is located at a motion boundary. Also let $\mathbf{d}_2(i)$ and $\mathbf{d}_1(i)$ denote the observed MVs of the ith block obtained using 16×16 and 8×8 matching block, respectively. We can see in Figure 3.3-(a) and (b) that $\mathbf{d}_2(i)$ sometimes fails to show the true motion of the ith block around motion boundaries. (See the MVs inside the black box.) This is because the 16×16 block used for obtaining $\mathbf{d}_2(i)$ in a BME algorithm includes objects having different movements as the ith block. Nevertheless, this false MV will have a large influence on the estimate $\mathbf{u}^t(i)$ since $\hat{c}_{2i}^{-1} \approx 0$ as seen in Figure 3.3-(c) and (d). In order to solve this problem, adjusting the influence of an observed MV $\mathbf{d}_k(i)$ to the estimation of the true MV $\mathbf{u}(i)$ is proposed by adaptively weighting the estimated noise variance $\hat{c}_{ki}^{-1}$ according to how suitable $\mathbf{d}_k(i)$ is for the target block.

Figure 3.3-(e) and (f) show the mean of absolute difference (MAD) of each 8×8 target block obtained using $\mathbf{d}_1$ and $\mathbf{d}_2$, respectively. Note that the MAD of the ith block

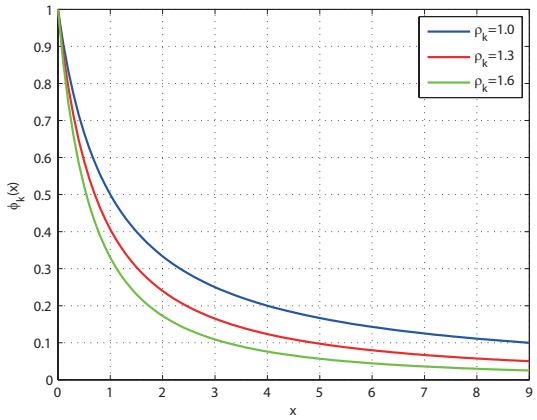


Figure 3.4: $\phi_k(x)$ with varying x and ρ_k.

when an MV $\mathbf{v}$ is applied is defined as $\epsilon_i(\mathbf{v}) = \frac{1}{|\mathcal{B}_i|}\sum_{\mathbf{m}\in\mathcal{B}_i}|I_c(\mathbf{m}) - I_p(\mathbf{m}+\mathbf{v})|$. We can see in the figures that the MADs obtained using $\mathbf{d}_2$ are normally higher than those obtained using $\mathbf{d}_1$ especially around the motion boundaries. This is because false MVs around motion boundaries generally bring about large block-matching errors. Therefore we can use $\epsilon_i(\mathbf{d}_k(i))$, the MAD of the ith block when the kth observation $\mathbf{d}_k(i)$ is applied, for adjusting the estimated noise variance $\hat{c}_{ki}^{-1}$, and finally propose using in (3.10)

$$c_{ki} = \omega_{ki}\cdot\hat{c}_{ki} = \hat{c}_{ki}\cdot\frac{\phi_k(\varepsilon_i(\mathbf{d}_k(i)))}{\sum_{k=1}^{M}\phi_k(\varepsilon_i(\mathbf{d}_k(i)))}, \tag{3.12}$$

where $k = 1, ..., M$, and $\phi_k(x) > 0$ is a decreasing function of $x \geq 0$ for the kth observation. It can be seen in (3.12) that if the observed MV results in large block-matching errors when applied to the target block, its contribution to the estimation of the true MV of the block is lessened by weighting the estimated noise variance of the observed MV.

The simple choice of the function $\phi_k(x)$ may be $\phi_k(x) = x^{-1}$ for $k = 1, ..., M$. In this dissertation, the following functions are proposed:

$$\phi_k(x) = (x+1)^{-\rho_k}, \;\; k = 1, ..., M, \tag{3.13}$$

where $\rho_k \geq 1$ is a constant satisfying the condition $\rho_{k+1} \geq \rho_k$ so $\phi_k(x) \geq \phi_{k+1}(x)$ holds for all $x > 0$. Figure 3.4 shows $\phi_k(x)$ with varying x and ρ_k. We can see in the figure that, given x, $\phi_k(x)$ decreases as ρ_k increases. The condition $\rho_{k+1} \geq \rho_k$, therefore, reflects that the contribution of the observed MV $\mathbf{d}_{k+1}(i)$ obtained from a larger block to the estimation of the true MV is lessened compared to that of the observed MV $\mathbf{d}_k(i)$ obtained from a smaller block even though both the observations result in the same MAD value. Note that such functions is proposed in order to preserve the details of the true MVF based on the fact that a smaller block, in general, is better than a larger block in presenting the details of the MVF [28]. Figure 3.3-(g) and (h), respectively, shows the weight ω_{1i} and ω_{2i} for adjusting the estimated noise variances in Figure 3.3-(c) and (d). We can find in the figures that ω_{1i} is normally higher than ω_{2i} especially around the motion boundaries. On the other hand, in non-motion boundaries, ω_{1i} and ω_{2i} show similar values so the reliable MVs of $\mathbf{d}_1$ and $\mathbf{d}_2$ can fairly contribute to the estimation according to their estimated noise variances.

Chapter 4

Single Hypothesis Bayesian Approach in a Bidirectional Framework

In this chapter, a method that simultaneously estimates the true forward and backward motion vector fields (MVFs) for two consecutive video frames is proposed. Under the assumption that the observed forward and backward MVFs, which are the results of a block-based motion estimation (BME), respectively are degraded versions of the true forward and backward MVFs by locally stationary additive Gaussian noise (AGN), the true MVFs are simultaneously estimated from the two observations by maximizing the joint posterior probability of them.

4.1 Problem Formulation

Given two adjacent frames I_p (the previous frame) and I_c (the current frame) of a video, let $\mathbf{d}_b = \{\mathbf{d}_b(i)|i \in \mathcal{I}\}$ denote the 2-D backward MVF obtained by a BME algorithm where $\mathbf{d}_b(i)$ ($\in \mathbb{R}^2$) is the backward MV of the i-th block of the frame I_c pointing to a matching block in I_p, $\mathcal{I}$ is the set of all block indices in a frame. Let's assume that $\mathbf{d}_b$, called the *observed backward MVF* or simply the *backward observation*, is a

degraded version of the true backward MVF $\mathbf{u}_b = \{\mathbf{u}_b(i)|i \in \mathcal{I}\}$ where $\mathbf{u}_b(i)$ ($\in \mathbb{R}^2$) is the true backward MV of the ith block. The degradation process is expressed as

$$\mathbf{d}_b = \varphi_b(\mathbf{u}_b) \tag{4.1}$$

for some degradation operation $\varphi_b(\cdot)$. As discussed in Chapter 2, the degradation operation $\varphi_b(\cdot)$ in (4.1) is modeled as locally stationary additive Gaussian noise (AGN) such that

$$\mathbf{d}_b(i) = \mathbf{u}_b(i) + \mathbf{e}_b(i), \quad i \in \mathcal{I}, \tag{4.2}$$

where $\mathbf{e}_b(i)$ denotes a realization of a 2-D random noise vector following locally stationary Gaussian with $\mathcal{N}(\mathbf{0}, c_i^b \cdot \mathbf{I})$. $\mathbf{0}$ and $\mathbf{I}$ are the 2×1 zero vector and the 2×2 identity matrix, respectively, and c_i^b is the noise variance. The *forward observation* $\mathbf{d}_f = \{\mathbf{d}_f(i)|i \in \mathcal{I}\}$ with $\mathbf{d}_f(i)$ ($\in \mathbb{R}^2$) denoting the forward MV of the i-th block of the frame I_p pointing to a matching block in I_c can be defined in the same manner with $\mathbf{d}_b$ using the observation model (4.2).

Under this assumption, the true backward MVF $\mathbf{u}_b$ and the true forward MVF $\mathbf{u}_f$ are estimated simultaneously from the observations $\mathbf{d}_b$ and $\mathbf{d}_f$ by maximizing the joint posterior probability

$$p(\mathbf{u}_b, \mathbf{u}_f \mid \mathbf{d}_b, \mathbf{d}_f, I_p, I_c) = \frac{p(\mathbf{d}_b, \mathbf{d}_f \mid \mathbf{u}_b, \mathbf{u}_f, I_p, I_c)p(\mathbf{u}_b, \mathbf{u}_f \mid I_p, I_c)p(I_p, I_c)}{p(\mathbf{d}_b, \mathbf{d}_f, I_p, I_c)}. \tag{4.3}$$

Therefore, the MAP estimates $\hat{\mathbf{u}}_b$ and $\hat{\mathbf{u}}_f$ can be obtained by

$$\{\hat{\mathbf{u}}_b, \hat{\mathbf{u}}_f\} = \underset{\{\mathbf{u}_b, \mathbf{u}_f\} \subset \mathcal{U}}{\operatorname{argmax}} [p(\mathbf{d}_b, \mathbf{d}_f \mid \mathbf{u}_b, \mathbf{u}_f, I_p, I_c)p(\mathbf{u}_b, \mathbf{u}_f \mid I_p, I_c)], \tag{4.4}$$

where $\mathcal{U}$ denotes the set of all candidate MVFs for I_c and I_p.

4.1.1 Observation likelihood

Let us first assume that $\mathbf{d}_b$ and $\mathbf{d}_f$ are conditionally independent given $\mathbf{u}_b$, $\mathbf{u}_f$, I_p, and I_c. This implies that the observed forward and backward MVFs are obtained by

a BME algorithm independently. Also, $\mathbf{d}_b$ and $\mathbf{d}_f$ respectively are independent of $\mathbf{u}_f$ and $\mathbf{u}_b$ so the joint probability $p(\mathbf{d}_b, \mathbf{d}_f \mid \mathbf{u}_b, \mathbf{u}_f, I_p, I_c)$ can be expressed as

$$p(\mathbf{d}_b, \mathbf{d}_f \mid \mathbf{u}_b, \mathbf{u}_f, I_p, I_c) = p(\mathbf{d}_b \mid \mathbf{u}_b, I_p, I_c) p(\mathbf{d}_f \mid \mathbf{u}_f, I_p, I_c). \tag{4.5}$$

Next, two assumptions are further made in calculating $p(\mathbf{d}_b \mid \mathbf{u}_b, I_p, I_c)$. Assumed are that the observations $\mathbf{d}_b(i)$, $i \in \mathcal{I}$, are conditionally independent given $\mathbf{u}_b$, I_p, and I_c, and that each $\mathbf{d}_b(i)$ is independent of $\mathbf{u}_b(j)$, $j \neq i$, $j \in \mathcal{I}$ so $p(\mathbf{d}_b(i)|\mathbf{u}_b, I_p, I_c) = p(\mathbf{d}_b(i)|\mathbf{u}_b(i), I_p, I_c)$. It then follows that $p(\mathbf{d}_b \mid \mathbf{u}_b, I_p, I_c) = \prod_{i\in\mathcal{I}} p(\mathbf{d}_b(i) \mid \mathbf{u}_b(i), I_p, I_c)$, and our observation model (4.2) finally leads us to

$$\begin{aligned} p(\mathbf{d}_b \mid \mathbf{u}_b, I_p, I_c) &= \prod_{i\in\mathcal{I}} p(\mathbf{d}_b(i) \mid \mathbf{u}_b(i), I_p, I_c) \\ &= \Big(\prod_{i\in\mathcal{I}} \frac{1}{2\pi c_i^b} \Big) \exp\Big(- \sum_{i\in\mathcal{I}} \frac{\|\mathbf{d}_b(i) - \mathbf{u}_b(i)\|^2}{2c_i^b} \Big). \end{aligned} \tag{4.6}$$

In the same manner, $p(\mathbf{d}_f \mid \mathbf{u}_f, I_p, I_c)$ can be expressed as

$$p(\mathbf{d}_f \mid \mathbf{u}_f, I_p, I_c) = \Big(\prod_{i\in\mathcal{I}} \frac{1}{2\pi c_i^f} \Big) \exp\Big(- \sum_{i\in\mathcal{I}} \frac{\|\mathbf{d}_f(i) - \mathbf{u}_f(i)\|^2}{2c_i^f} \Big). \tag{4.7}$$

Note that the noise variances c_i^b and c_i^f correspond to our confidence in $\mathbf{d}_b(i)$ and $\mathbf{d}_f(i)$, respectively, and are estimated from the observations via maximum likelihood estimation as described in chapter 2.

4.1.2 Joint prior distribution of true motion vector fields

It is assumed that the joint prior distribution of the true forward and backward MVFs $p(\mathbf{u}_b, \mathbf{u}_f|I_c, I_p)$ follows the distribution

$$p(\mathbf{u}_b, \mathbf{u}_f|I_p, I_c) = \frac{1}{Z} \exp\Big(- U(\mathbf{u}_b, \mathbf{u}_f|I_c, I_p) \Big), \tag{4.8}$$

where Z is a normalizing constant. The energy function $U(\mathbf{u}_b, \mathbf{u}_f|I_c, I_p)$ is defined as

$$\begin{aligned} U(\mathbf{u}_b, \mathbf{u}_f|I_c, I_p) = \sum_{i\in\mathcal{I}} \Big[\gamma \cdot \sum_{j\in\mathcal{N}_i} \Big(V_{ij}^1(\mathbf{u}_b) + V_{ij}^1(\mathbf{u}_f) \Big) \\ + \zeta \cdot \Big(V_i^2(\mathbf{u}_b, \bar{\mathbf{u}}_b) + V_i^2(\mathbf{u}_f, \bar{\mathbf{u}}_f) \Big) \Big], \end{aligned} \tag{4.9}$$

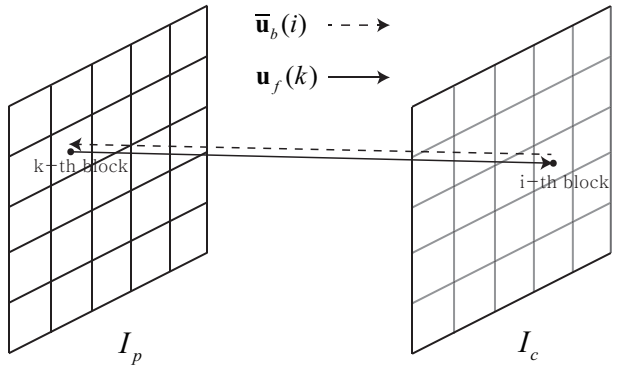


Figure 4.1: The dual MV $\bar{\mathbf{u}}_b(i) = -\mathbf{u}_f(k)$ of $\mathbf{u}_b(i)$.

where $\mathcal{N}_i$ is the index set of the 8 blocks neighboring the i-th block. $\bar{\mathbf{u}}_b(i)$ ($\bar{\mathbf{u}}_f(i)$), called the *dual* MV of $\mathbf{u}_b(i)$ ($\mathbf{u}_f(i)$), is obtained by reversing the forward (backward) MV in $\mathbf{u}_f$ ($\mathbf{u}_b$) that points out the i-th block of I_c (I_p). Figure 4.1 shows the dual MV $\bar{\mathbf{u}}_b(i)$ of $\mathbf{u}_b(i)$. $V_{ij}^1(\mathbf{u})$ is the anisotropic potential function proposed in chapter 2 and has the following form:

$$V_{ij}^1(\mathbf{u}) = \mu_{ij}||\mathbf{u}(i) - \mathbf{u}(j)||^2. \tag{4.10}$$

As described in chapter 2, the weight factor μ_{ij}, $i \in \mathcal{I}$, $j \in \mathcal{N}_i$ controls the influence of the neighboring MV $\mathbf{u}(j)$ on the center MV $\mathbf{u}(i)$ during the refinement process. It shows a large value when the neighboring MV is most likely to show the true motion of the center block. Otherwise, it shows a small value.

The potential function $V_i^2(\mathbf{u}, \bar{\mathbf{u}})$, which is devised to reflect the fact that the true backward (forward) MV should be consistent with the reverse of the true forward (backward) MV that points out the corresponding block of the true backward (forward) MV, is defined as

$$V_i^2(\mathbf{u}, \bar{\mathbf{u}}) = \begin{cases} \min\Big(||\mathbf{u}(i) - \bar{\mathbf{u}}(i)||^2, l\Big), & \text{if } \bar{\mathbf{u}}(i) \text{ exists} \\ l, & \text{otherwise,} \end{cases} \tag{4.11}$$

where ζ and l are positive constants. It is seen in (4.11) that the potential V_i^2 is pro-

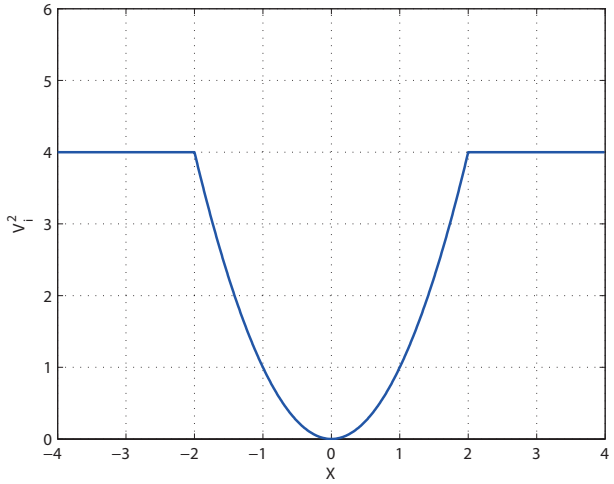


Figure 4.2: Example of the potential function V_i^2 for the 1-dimensional MVs $u(i)$ and $\bar{u}(i)$. Note that $V_i^2 = \min(\{X, l\})$ where $X = |u(i) - \bar{u}(i)|^2$, $l = 4$.

portional to the distance between $\mathbf{u}(i)$ and $\bar{\mathbf{u}}(i)$ so, in order to decrease the potential, $\mathbf{u}(i)$ should be close to $\bar{\mathbf{u}}(i)$. If $\bar{\mathbf{u}}(i)$, however, does not present the true motion of the i-th block, the influence of $\bar{\mathbf{u}}(i)$ to $\mathbf{u}(i)$ should be lessened. It is done by switching off the consistency constraint $||\mathbf{u}(i) - \bar{\mathbf{u}}(i)||^2$ if the distance between $\mathbf{u}(i)$ and $\bar{\mathbf{u}}(i)$ is greater than l. This means that it is better to pay l for the potential if $\mathbf{u}(i)$ is quite different from $\bar{\mathbf{u}}(i)$ so $\mathbf{u}(i)$ is influenced by only its spatially neighboring MVs. Figure 4.2 shows an example of the potential function V_i^2 for the 1-dimensional MVs $u(i)$ and $\bar{u}(i)$ when $l = 4$. Note that the potential function has been used as the piecewise smoothness constraint in image denoising [52].

4.2 MAP Solution

Now the MAP estimates are obtained by combining (4.5) and (4.8) in (4.4), and consequently by minimizing of the total energy given by

$$\mathcal{E}(\mathbf{u}_b, \mathbf{u}_f) = U(\mathbf{d}_b, \mathbf{d}_f | \mathbf{u}_b, \mathbf{u}_f, I_c, I_p) + U(\mathbf{u}_b, \mathbf{u}_f | I_c, I_p), \tag{4.12}$$

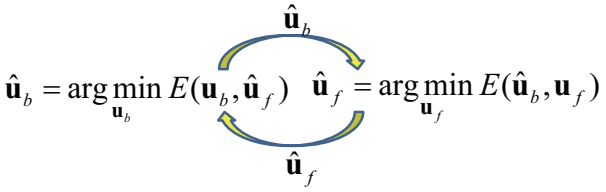


Figure 4.3: The proposed alternation method.

where

$$U(\mathbf{d}_b, \mathbf{d}_f | \mathbf{u}_b, \mathbf{u}_f, I_c, I_p) = \sum_{i \in \mathcal{I}} \Big(\frac{\|\mathbf{d}_b(i) - \mathbf{u}_b(i)\|^2}{2c_i^b} + \frac{\|\mathbf{d}_f(i) - \mathbf{u}_f(i)\|^2}{2c_i^f} \Big). \quad (4.13)$$

It is nontrivial to simultaneously find $\mathbf{u}_b$ and $\mathbf{u}_f$ that minimize the total energy $\mathcal{E}(\mathbf{u}_b, \mathbf{u}_f)$. In this dissertation, an alternation method is proposed for minimizing the energy in (4.12) as follows: Given the current estimate $\hat{\mathbf{u}}_b$, $\hat{\mathbf{u}}_f$ that minimizes $\mathcal{E}(\hat{\mathbf{u}}_b, \mathbf{u}_f)$ is first found. Next, $\hat{\mathbf{u}}_b$ is updated with the $\hat{\mathbf{u}}_f$ by minimizing $\mathcal{E}(\mathbf{u}_b, \hat{\mathbf{u}}_f)$, and $\hat{\mathbf{u}}_b$ is again used for renewing $\hat{\mathbf{u}}_f$. The repeated updating of $\hat{\mathbf{u}}_f$ and $\hat{\mathbf{u}}_b$ continues in iterations. Figure 4.3 shows the proposed alternation method.

The iterative conditional mode (ICM) algorithm [47] is used to minimize the total energy $\mathcal{E}(\mathbf{u}_b, \hat{\mathbf{u}}_f)$. The algorithm sequentially updates the kth estimate into the (k+1)th estimate by minimizing for each i the total potential $V_T(\mathbf{u}_b(i), \hat{\mathbf{u}}_f(i))$ where $\mathcal{E}(\mathbf{u}_b, \hat{\mathbf{u}}_f) = \sum_{i \in \mathcal{I}} V_T(\mathbf{u}_b(i), \hat{\mathbf{u}}_f(i))$. Note that the observations $\mathbf{d}_b$ and $\mathbf{d}_f$ are used as the initial estimates of $\mathbf{u}_b$ and $\mathbf{u}_f$, respectively.

Chapter 5

Experimental Results

In this chapter, the proposed algorithms are applied to motion-compensated frame rate up conversion (MC-FRUC) to demonstrate their performances.

5.1 Experimental Settings

The proposed methods are compared with nine existing algorithms. Among them are four motion refinement algorithms: the vector regularization method for images with periodic patterns (VRP) [40], the error-weighted vector regularization algorithm (EVR) [39], the conventional vector regularization method (CVR) [37], the correlation-based multistage MVF refinement algorithm (CMMR) [43]; and five motion estimation algorithms: the multi-frame MAP algorithm (MFMAP) [21], progressively reduced block-size motion estimation algorithm (PRBME) [31], block-based motion estimation algorithm with block-overlap cost (BMEBC) [22], hierarchical BME algorithm with reliable motion adaption (HBMERM) [28], and iterative motion estimation algorithm using duality of MVF (IterME) [33].

The experiments use 20 test videos, *Foreman*, *Coastguard*, *Silent*, *Mobile*, *Hall*, *City*, *Ice*, *Flower*, *Highway*, *Harbour*, *WalkingLion*, *Soccer*, *Drama2*, *Runtrack*, *Subtitle*,

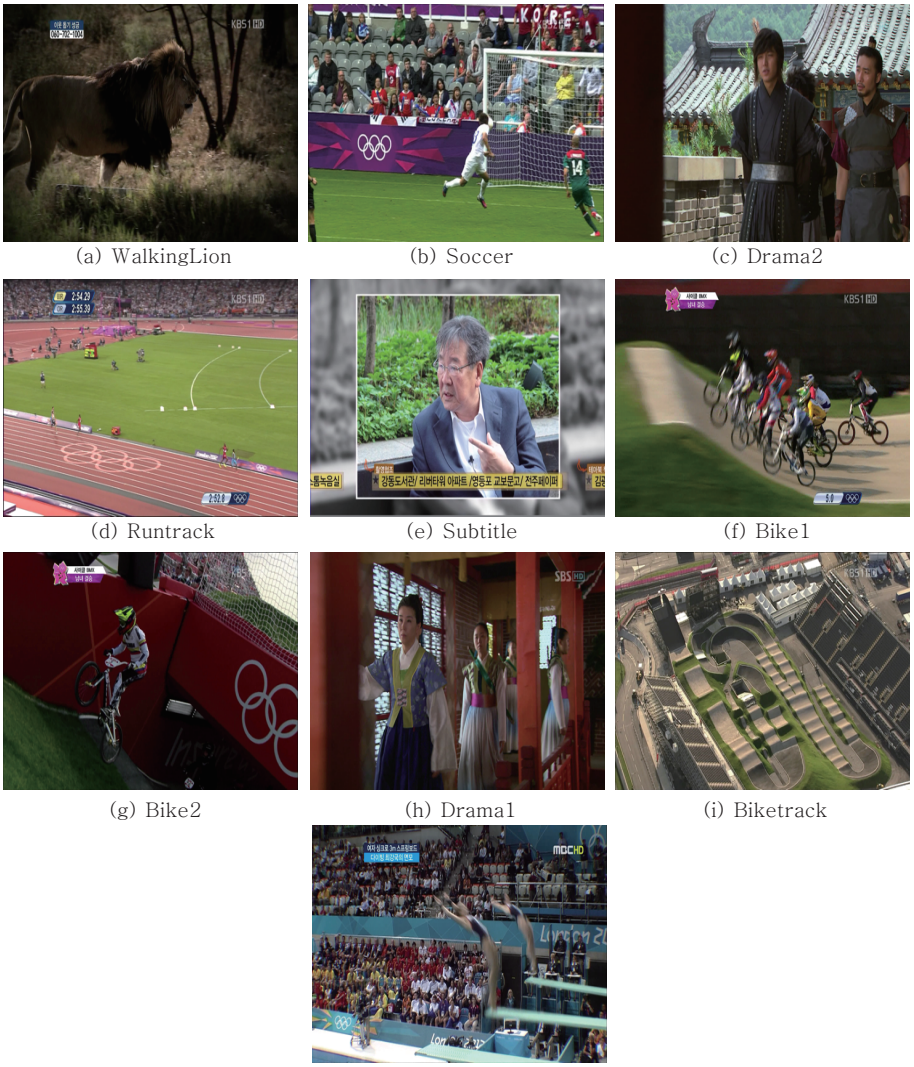
(a) Foreman (b) Coastguard (c) Ice

(d) Silent (e) Mobile (f) Flower

(g) Hall (h) City (i) Harbour

(j) Highway

Figure 5.1: Representative frames of CIF test video sequences.



(a) WalkingLion (b) Soccer (c) Drama2

(d) Runtrack (e) Subtitle (f) Bike1

(g) Bike2 (h) Drama1 (i) Biketrack

(j) Diving

Figure 5.2: Representative frames of qFHD test video sequences.

Bike1, *Bike2*, *Drama1*, *Diving*, and *Biketrack*. Figure 5.1 and 5.2 show the representative frames of the test video sequences. The first ten videos are composed of 240∼300 frames of size 352×288 pixels (CIF) at 30 fps. The rest have 100∼300 frames of size 960×544 pixels (qFHD) at 60 fps. For the experiments, the observed forward and backward MVFs between two successive odd frames of the test video sequences are obtained by the full search block-matching algorithm (BMA). For the first ten videos, the BMA is utilized with the block size 8×8 pixels over the search range +/-24 pixels both in horizontal and vertical directions; for the rest, the algorithm with the block size 8×8 pixels over the search range +/-32 is utilized to be able to accommodate, for the videos of qFHD size, large motions. Block matching is performed at integer pixel accuracy and with the sum of absolute difference (SAD) as the matching criterion. The observed MVFs are also used as the initial MVFs in implementing MFMAP and BMEBC. For PRBME and HBMERM, the block size 32×32, 16×16, and 8×8 are used for the first, second, and the third level, respectively, and the frame interpolation described in section 5.2 uses 8×8 pixel blocks. Note that the full-search BMA with the block sizes 32×32, 16×16, and 8×8 pixels are used for MHBA to obtain the multiple observed MVFs.

The parameter γ in (2.13) and (4.9) and ζ in (4.9) determine the overall smoothing strength of the proposed algorithms. It is found from experiments that the optimal values of the parameters depend on the reliability of the observed MVF and the scene characteristics of the video frames. For example, small values of the parameters are preferred when the observed MVF seems reliable and objects in frames show complex complex motions. In this dissertation, $\gamma = 4.0$, $\gamma = 2.5$, and $\gamma = 4.0$ respectively are used for SHBA, MHBA, and SHBABF since they give satisfactory estimation results on the test video sequences. ρ_k's in (3.13) are set $\rho_1 = 1.0$, $\rho_2 = 1.25$, and $\rho_3 = 1.5$ respectively for the observed MVFs $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$. Note that $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$ respectively are obtained by the full-search BMA with 8×8, 16×16, and 32×32 pixel

blocks. l and ζ in (4.11) are set $l = 16$ and $\zeta = 4.0$. Finally, the parameters for the other algorithms are set as the corresponding papers suggested.

5.2 Performance Evaluation

In this section, the proposed algorithm is compared with the nine algorithms both subjectively and objectively. For the subjective comparisons, the original even frame and its interpolation results are presented since the true MVF of the test videos are unknown. The interpolated frames are obtained from the estimated backward MVFs by means of overlapped-block motion compensation (OBMC) such that

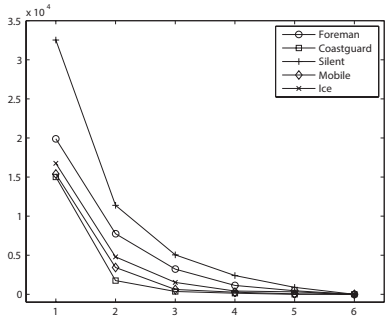
$$I_{intp}\Big(\mathbf{m} + \frac{1}{2}\mathbf{v}\Big) = \tau_1 \cdot I_p(\mathbf{m} + \mathbf{v}) + \tau_2 \cdot I_c(\mathbf{m}), \tag{5.1}$$

where $\tau_1 = \tau_2 = \frac{1}{2}$, the block size for motion compensation is 8×8 pixels, and the block overlaps are 2 pixel wide. If no MV is available to perform interpolation, i.e., at "holes", the MVs are first obtained from the neighbors by VMF, and (5.1) is used with $\tau_1 = 1$ and $\tau_2 = 0$. This is because the holes created by a backward MVF correspond usually to covered areas. For CMMR, bidirectional MCFI is used because the algorithm estimates the true MVF of the to-be-interpolated frame. (For more details, see [43], p. 741.)

For objective comparisons, the PSNR and SSIM [49] of the interpolated even frames are presented in the tables presented in subsequent sections. The values in the tables are computed first for each even numbered frame and then averaged over the whole length of the video.

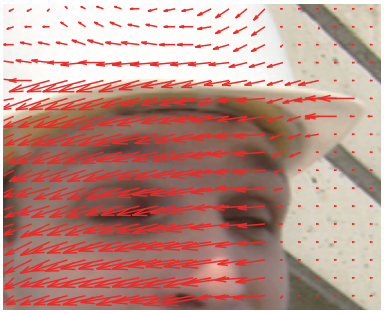
5.2.1 Performance of SHBA

The convergence of SHBA is first checked. The backward MVF from 19-th to 17-th frame of the CIF test videos are used to show in Figure 5.3 how the energy $\mathcal{E}(\mathbf{u}^k)$ in (2.23) varies with the iteration number k, where the energy is shifted by a constant for

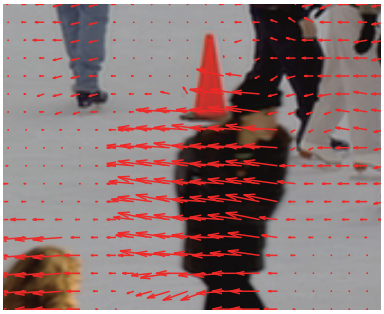


(a)

(b) *Coastguard*



(c) *Foreman*



(d) *Ice*

Figure 5.3: (a) the total energy $\mathcal{E}(\mathbf{u}^k)$, shifted by a constant, versus iteration number k. (b)$\sim$(d) MVF smoothing results.

better presentation. It is seen that the energy $\mathcal{E}(\mathbf{u}^k)$ quickly converges to a minimum value in only 5$\sim$6 iterations. Figure 5.3-(b)$\sim$(d) show the MAP estimates of the true backward MVFs when $k = 6$. The figures also verify that SHBA well smoothes MVFs while preserving motion boundaries in only few iterations.

Next it is shown that how the diagonals (η_i and κ_i) of the covariance matrix C_i in (2.25) affect the estimation of the true MVF. Figure 5.4 shows the estimated MVFs and the frames interpolated using the MVFs. The MVFs in Figure 5.4-(c) and (d) are the estimation results of the true MVF in the red box of Figure 5.4-(a). To obtain Figure 5.4-(c), C_i is estimated from $\mathcal{N}_i^{\mathbf{d}}$ while Figure 5.4-(d) uses $\mathcal{N}_i$, and Figure 5.4-(g) and (h) visualize $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ used for Figure 5.4-(c) and (d), respectively. We can see

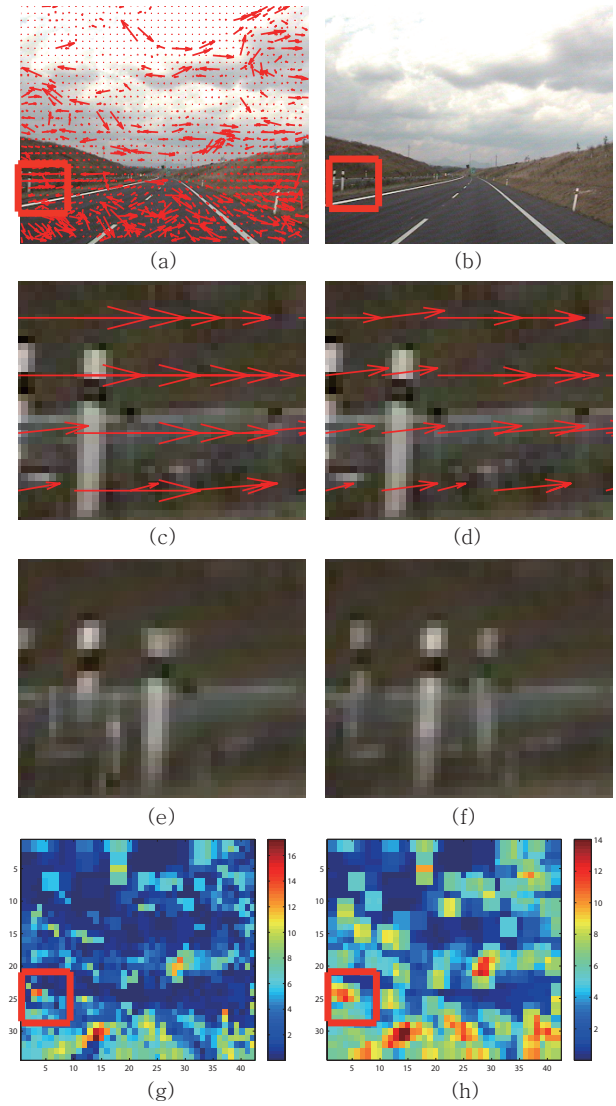


Figure 5.4: (a) the observed backward MVF of the 195-th frame of *Highway*, (b) target frame, the 194-th frame of *Highway*, (c) and (d) the MVFs estimated using $\hat{C}_i$ obtained from $\mathcal{N}_i^{\mathbf{d}}$ and $\mathcal{N}_i$, respectively, (e) and (f) frames interpolated using the MVFs in (c) and (d), respectively. (g) and (h) respectively visualize $\sqrt{\frac{1}{2}trace\{\hat{C}_i\}}$ values obtained from $\mathcal{N}_i^{\mathbf{d}}$ and $\mathcal{N}_i$.

(a)

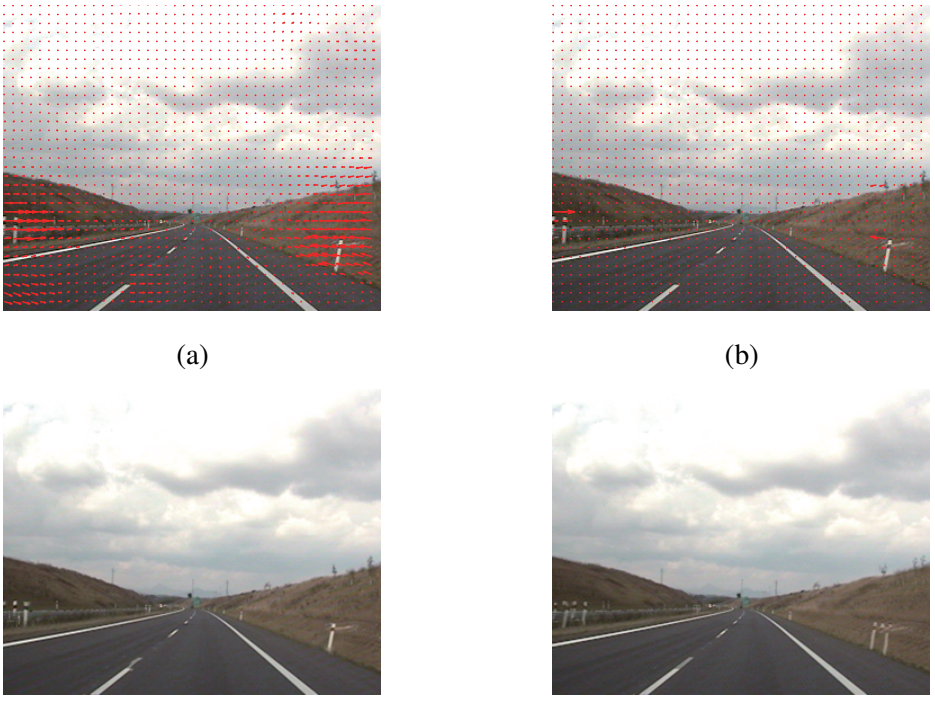
(b)

Figure 5.5: (a) the whole of the frame interpolated using the MVF of Figure 5.4-(c), (b) the whole of the frame interpolated using the MVF of Figure 5.4-(d).

in Figure 5.4-(e) and (f), which are the frames interpolated using the MVFs in Figure 5.4-(c) and (d), respectively, that $\hat{C}_i$ from $\mathcal{N}_i^{\mathbf{d}}$ stops over-smoothing better than that from $\mathcal{N}_i$. On the other hand, it is seen in Figure 5.5 that $\hat{C}_i$ from $\mathcal{N}_i^{\mathbf{d}}$ also correctly identifies false MVs in the observed MVF of Figure 5.4-(a) so the false MVs are well corrected during the estimation.

To show how the weight factor μ_{ij} in (2.25) affect the estimation, in Figure 5.6, the estimated true MVFs of Figure 5.4-(a) and the frames interpolated using the MVFs are presented. Note that C_i is estimated from $\mathcal{N}_i^{\mathbf{d}}$. The MVFs in Figure 5.6-(a) and (b) are obtained under $\mu_{ij} = (1-\zeta)\alpha_{ij} + \zeta\beta_{ij}$ and $\mu_{ij} = \frac{1}{|\mathcal{N}_i|}$, respectively. Here, $\mu_{ij} = \frac{1}{|\mathcal{N}_i|}$ means that the neighboring MVs are averaged with equal weights during the estimation. Figure 5.6-(c) and (d) show the frames interpolated using the MVFs of Figure 5.6-(a) and (b), respectively. We can see in the figure that when the neighboring MVs are averaged with equal weights, false MVs in the observation are improperly affected by their neighbors.

Figure 5.7 and 5.8 show how correctly the proposed algorithm smoothes false MVs of the observed MVFs. In Figure 5.7, the man is turning his head to his left so the motion boundary corresponds to the outline of the head. It is seen that SHBA as



(a) (b)

(c) (d)

Figure 5.6: (a) the MVF estimated under $\mu_{ij} = (1-\zeta)\alpha_{ij} + \zeta\beta_{ij}$ using the observed MVF in Figure 5.4-(a), (b) the MVF estimated under $\mu_{ij} = \frac{1}{|\mathcal{N}_i|}$ using the observed MVF in Figure 5.4-(a), (c) the frame interpolated using the MVF of Figure 5.6-(a), (d) the frame interpolated using the MVF of Figure 5.6-(b).

well as BMEBC and MFMAP well preserve the edge of the safety helmet and the collar. In addition the algorithms show satisfactory results in smoothing MVs of the face. The superiority of SHBA can be observed in Figure 5.8. The figure shows the two cut-outs from the interpolated frames where the subtitles move from right to left. It is found that that SHBA preserves the subtitles better than the other algorithms. On the other hand, the MVF regularization algorithms, VRP and EVR show serious weaknesses because of their equal or false treatment of the neighboring MVs during their smoothing process.

Figure 5.7: Original and its interpolated 130-th frames of *Foreman*. (a) Original, (b) Proposed ($k = 6$), (c) BMEBC ($k = 3$), (d) VRP ($k = 6$), (e) EVR ($k = 6$), (f) CMMR, (g) MFMAP ($k = 3$), (h) PRBME.

(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 5.8: Original and its interpolated 60-th frames of *Subtitle*. (a) Original, (b) Proposed ($k = 6$), (c) BMEBC ($k = 3$), (d) VRP ($k = 6$), (e) EVR ($k = 6$), (f) CMMR, (g) MFMAP ($k = 3$), (h) PRBME.

To demonstrate the performance of SHBA on images with periodic patterns, *City* and *Soccer* test videos are used in Figure 5.9 and 5.10, respectively. In the figures, the buildings and the goal net slowly move from right to left due to the panning of the camera. The figures show that SHBA as well as BMEBC outperform the rest in the periodic patterned areas. CMMR also fairly well refines false motions through the block merging process, but it sometimes fails and results in clusters of false motions. (See the building and the net near the man's head.) This indicates that the merging method is not immune to choosing wrong vectors. By contrast the inferior performance of CVR is very noticeable. The matching error surface of a block in the building and the goal net has multiple prominent peaks at false motions, which confuse CVR as it uses the sharpness of the peak as the reliability measure. Multiple prominent peaks on the matching error surface also invalidate the minimum matching error value itself as a reliability measure as seen in the results of MFMAP.

The average PSNR and SSIM performances of SHBA and the existing algorithms are presented in Figure 5.11. The values in the figure are the total average of the PSNRs of the interpolated frames for the twenty test videos. (For more details, see Table 1$\sim$9.) They show that SHBA outperforms the existing methods, competing with state-of-the-art BME algorithms, BMEBC and MFMAP. In Table 5.5, the effect of making δ variable in the threshold of (2.11) as compared to a constant δ is tested. The constant value of δ is chosen such that it maximizes the average PSNR of the interpolated frames and naturally differs from video to video. It should be noted that this value is fine-tuned for the test to each video, but which is not possible in practice.

The performance variation of SHBA against the quality of the observed MVFs is now presented. Figure 5.12 shows how the total average of PSNR performance for the ten CIF test videos varies against the search range used for obtaining the observed MVFs. We can see in the figure that the performance of SHBA decreases as the search range increases. However, the rate of the decrease is very small and comparable to

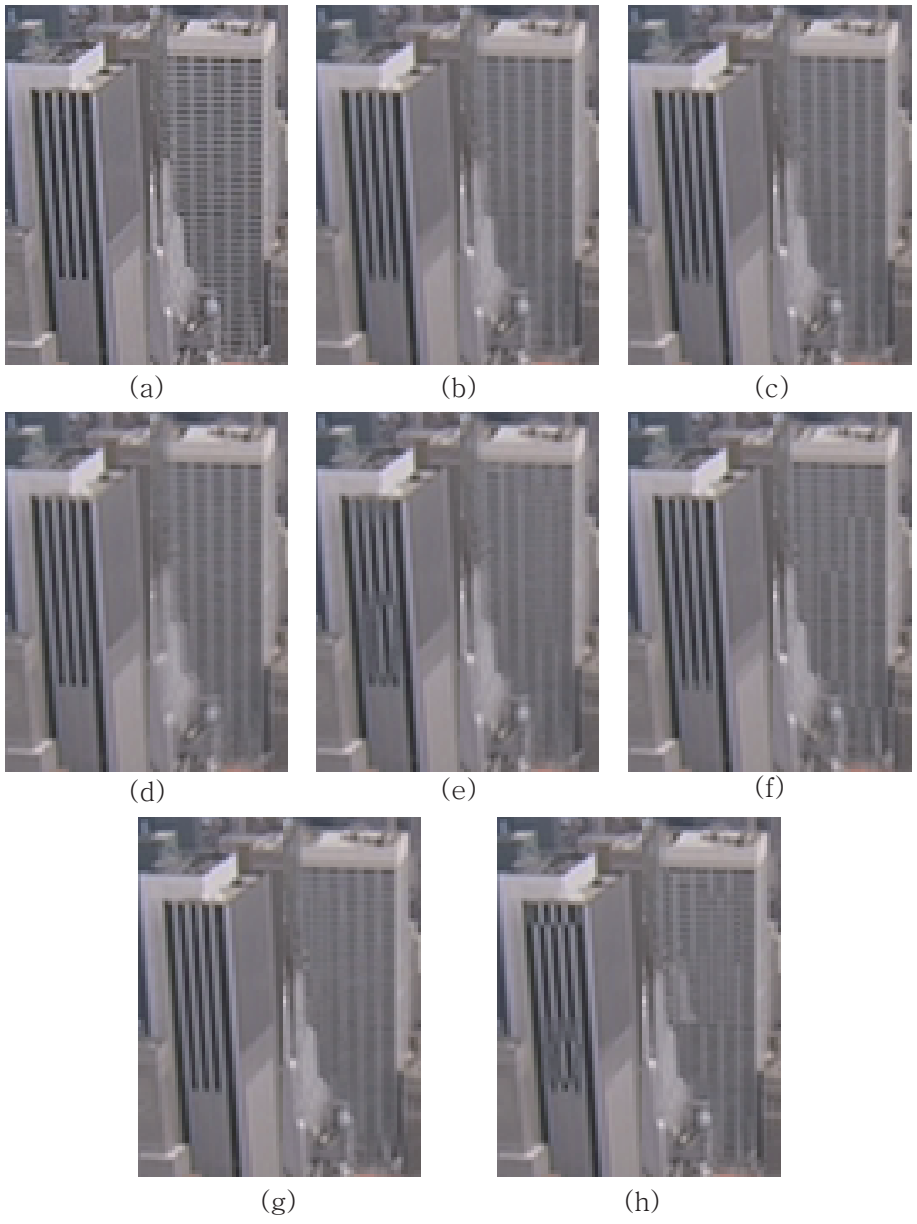


Figure 5.9: Original and its interpolated 52-th frames of *City*. (a) Original, (b) Proposed ($k = 6$), (c) BMEBC ($k = 3$), (d) VRP ($k = 6$), (e) EVR ($k = 6$), (f) CMMR, (g) MFMAP ($k = 3$), (h) PRBME.

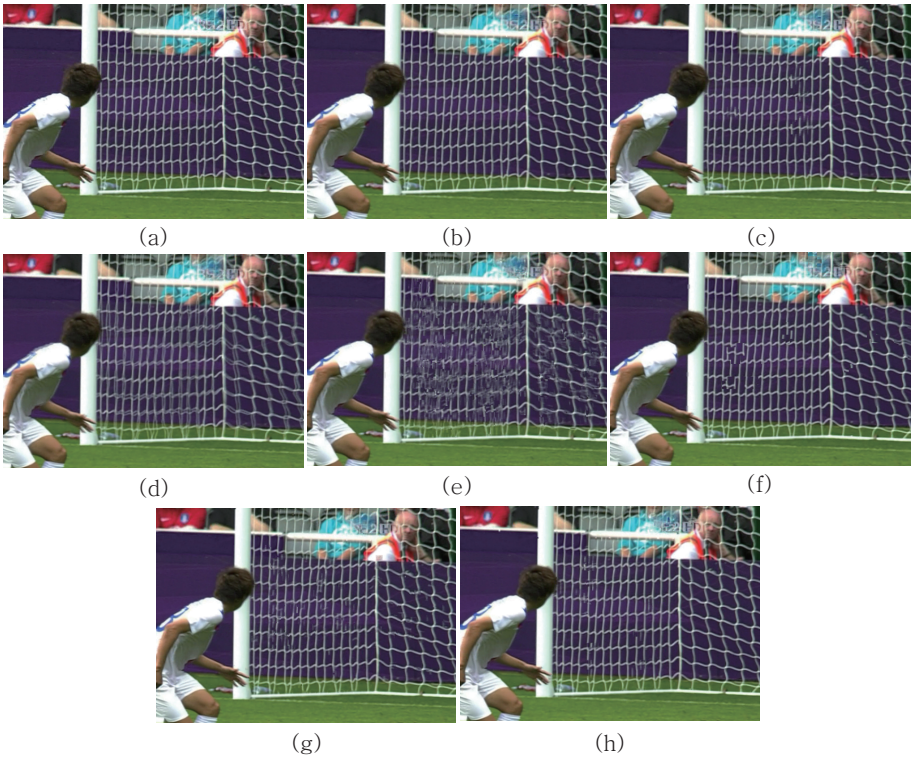


Figure 5.10: Original and its interpolated 188-th frames of *Soccer*. (a) Original, (b) Proposed ($k = 6$), (c) BMEBC ($k = 3$), (d) VRP ($k = 6$), (e) EVR ($k = 6$), (f) CMMR, (g) MFMAP ($k = 3$), (h) PRBME.

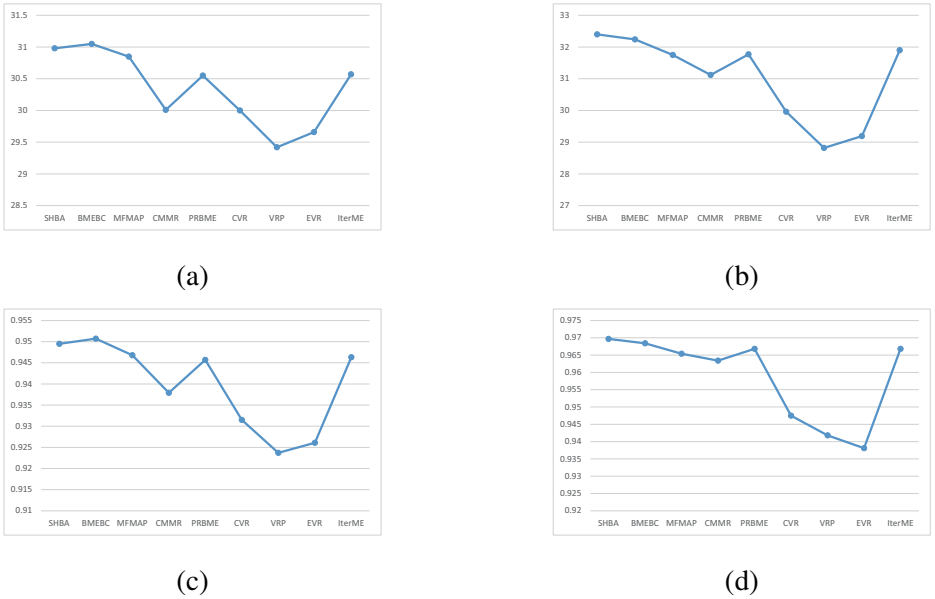


Figure 5.11: Total average PSNR performances for (a) the whole CIF videos and (b) the whole qFHD videos. Total average SSIM performances for (c) the whole CIF videos and (d) the whole qFHD videos.

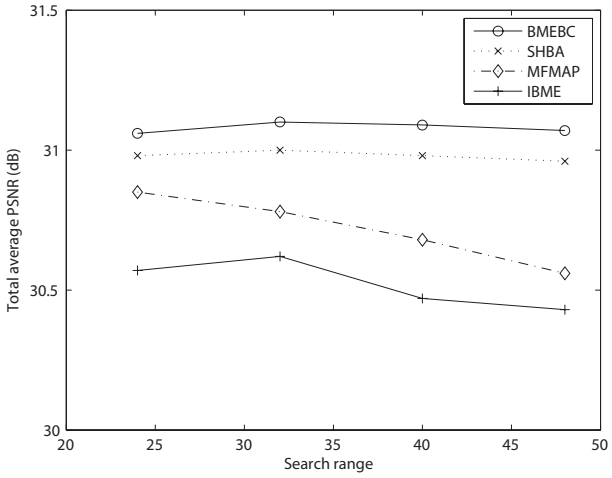


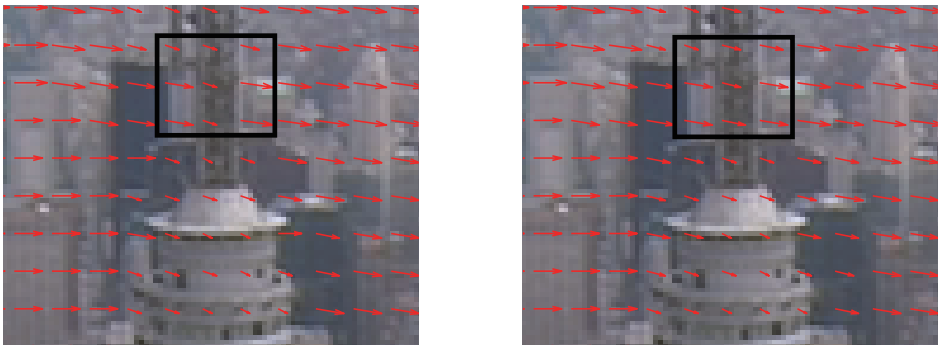
Figure 5.12: Performance variation of SHBA against the search range.

BMEBC.

Finally, the average computing times of the algorithms are compared. Each algorithm is implemented on a personal computer (OS: Windows 7, CPU: Intel Core i7 @3.20GHz, RAM: 12GB) using Matlab R2011b, and the execution times are summarized in Table 5.10. It is seen in the table that VRP runs fastest while both MFMAP and BMEBC take longest. Calculation of the SAD for all candidate MVs in the search range accounts for a great part of the run times of MFMAP, BMEBC, and IterME. SHBA is much faster than the two algorithms since it calculates the SAD only for the neighboring MVs. In particular, in an iteration for each block, our method needs 8 SAD calculations for the 8 neighboring MVs, while the first two need $(2W+1)^2$ (=2401) when the search range is +/-W (=24) in both directions.

5.2.2 Performance of MHBA

It is first demonstrated how well the weight ω_{ki} proposed in (3.12) works. Figure 5.13 shows the estimated true backward MVF of the test video sequence in Figure 3.3. To obtain the result in Figure 5.13-(a), ω_{ki} is calculated according to (3.12). On the other hand, the result in Figure 5.13-(b) is obtained by letting $\omega_{ki} = \frac{1}{M}$ where M is the number of the observations. It is be seen in the figure that the motion boundaries are well preserved when the weight ω_{ki} is calculated according to (3.12). In contrast, the motion boundaries are corrupted when $\omega_{ki} = \frac{1}{M}$. (See the MVs inside the black box.) To demonstrate the performance of MHBA on images with periodic patterns, *Mobile* test video is used in Figure 5.14. In the figures, the calendar has almost the zero motion and the red ball moves left from right. It is seen in the figure that MHBA and BMEBC outperform the rest. HBMERM sometimes fails to producing smooth MVs in the calendar even if it utilizes the multiple observed MVFs obtained from blocks of different sizes. Figure 5.15 shows how well MHBA smoothes MVFs while preserving motion boundaries. In figure 5.15, a white boat moves from left to right,



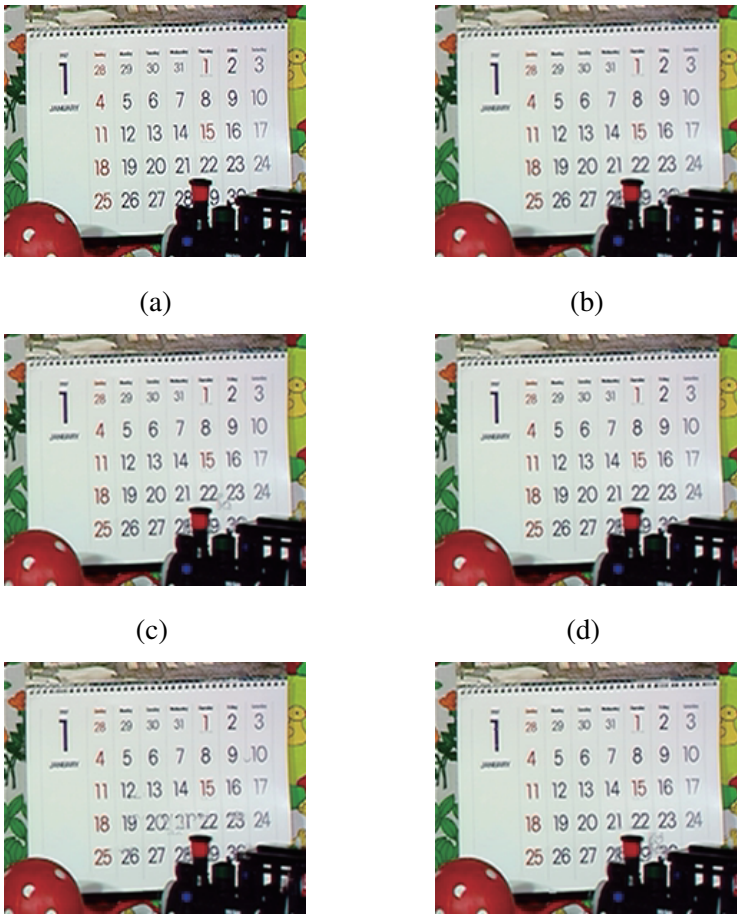
(a) (b)

Figure 5.13: (a) estimation result with ω_{ki} proposed in (3.13), (b) estimation result with $\omega_{ki} = \frac{1}{M}$.

therefore, motion boundaries correspond to the outline of the boat. It is seen in the figure that MHBA well preserves motion boundaries even if it utilizes the multiple observed MVFs obtained from blocks of different sizes. This is consequence that the influence of the observed MVs, which are obtained from larger matching blocks, is well suppressed by the proposed weight factor ω_{ki} in (3.12). PSNR and SSIM performances of MHBA are compared with those of the existing methods in Figure 5.16. (Fore more details, see Table 1∼9.) The figure verifies that MHBA outperforms the existing methods. MHBA especially shows its superiority on test videos containing frames with periodic patterned areas (*Mobile*, *City*, *Soccer*, *Runtrack*, and *Biketrack*).

5.2.3 Performance of SHBA-BF

In this section, the performance of SHBA-BF is compared with SHBA and MHBA. Figure 5.16 shows the average PSNR and SSIM performances of the proposed three algorithms for the twenty test videos. It is seen in the figure that MHBA shows the best performance while SHBA shows the worst performance among the three algorithm. SHBA-BF outperforms SHBA in almost all the test video sequences. Figure

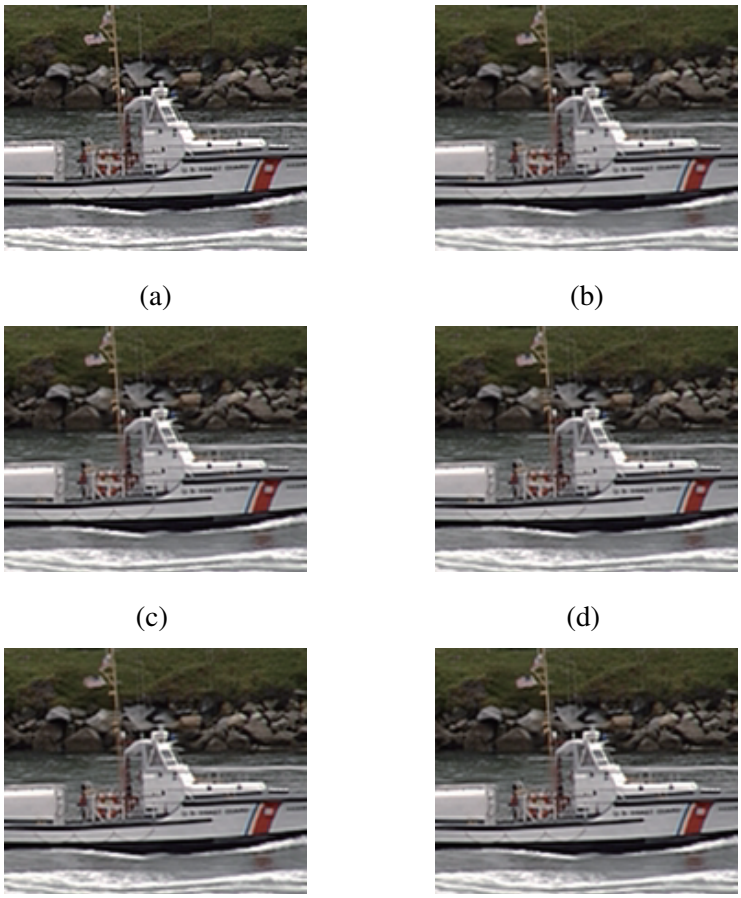


(a) (b)

(c) (d)

(e) (f)

Figure 5.14: Original and its interpolated 48-th frames of *Mobile*. (a) Original, (b) MHBA ($k = 6$), (c) SHBA ($k = 6$), (d) BMEBC ($k = 3$), (e) MFMAP ($k = 3$), (f) HBMERM.

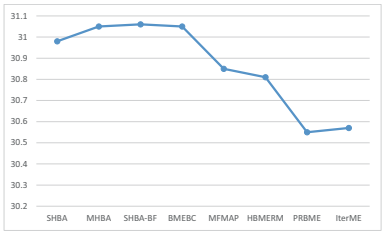


(a) (b)

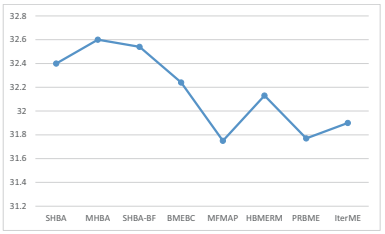
(c) (d)

(e) (f)

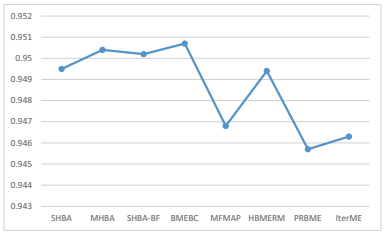
Figure 5.15: Original and its interpolated 220-th frames of *Coastguard*. (a) Original, (b)MHBA ($k = 6$), (c) SHBA ($k = 6$), (d) BMEBC ($k = 3$), (e) MFMAP ($k = 3$), (f) HBMERM.



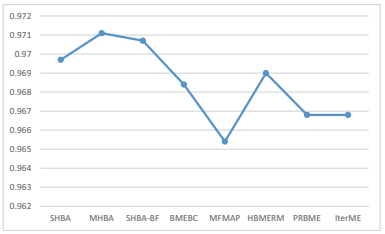
(a)



(b)



(c)

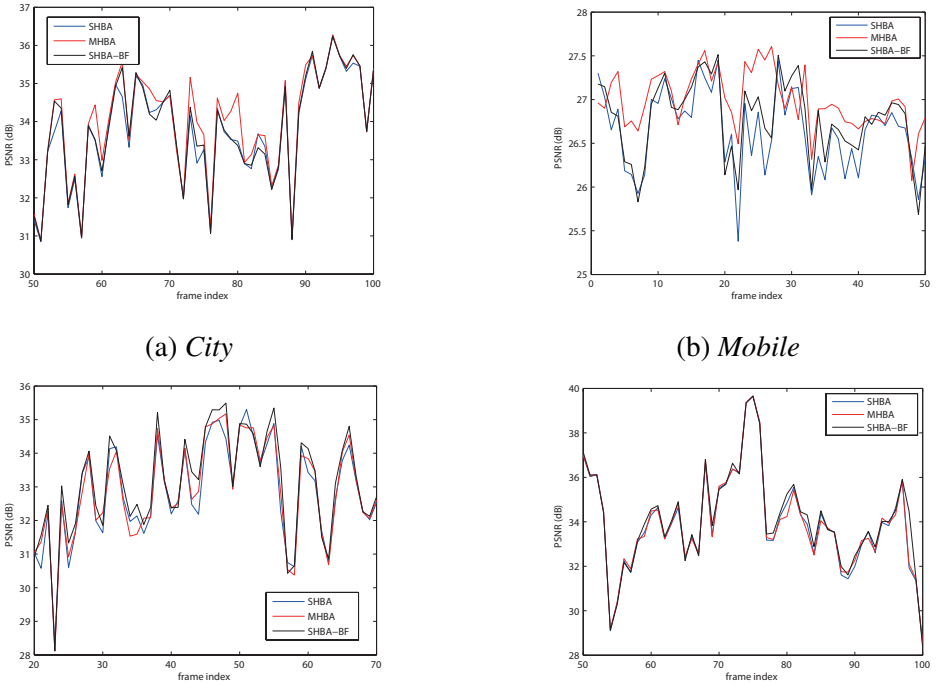


(d)

Figure 5.16: (a) average PSNR performances for the whole CIF videos, (b) average PSNR performances for the whole qFHD videos, (c) average SSIM performances for the whole CIF videos, (d) average SSIM performances for the whole qFHD videos.

5.17 compares PSNR performances of the proposed three algorithms for the four test video sequences. As seen in the figure, the performance improvement owing to the utilization of multiple observed MVFs is noticeable especially on the test videos containing frames with periodic patterned areas. (*Mobile*, *City*.) In contrast, the performance improvement owing to the simultaneous estimation of the true forward and backward MVFs is noticeable on the test videos containing objects with fast and dynamic motion characteristics. (*Ice*, *Silent*.) Figure 5.18 supports this fact. In the figure, the average PSNR performances of the proposed algorithms as well as the existing methods are presented for *Drama1* and *Bike1* test videos. In *Drama1*, some women walk quickly from right to left. On the other hand, in *Bike1*, some people ride motorbikes fast and the camera follows them. In order to capture large movements of the objects in the videos, the proposed algorithms as well as BMEBC use the observations obtained from full search block-matching with block size 8×8 pixels over the search range +/-64. For IBME, the search range is set to +/-64. We can see in the figure that SHBA-BF shows the best performance among the five algorithms.

Finally, in order to show how accurately the proposed algorithms estimate the true MVF, the average estimation errors are presented in Figure 5.19 for the test videos in Figure 2.1-(a) (*Hall* video), Figure 2.9-(a) (*Foreman* video), Figure 2.9-(b) (*Coastguard* video), and Figure 2.9-(d) (*City* video). The average estimation error is obtained by averaging the l_1 norm of the estimation errors, i.e., $||\hat{\mathbf{u}}(i) - \mathbf{u}(i)||_1$ where $\mathbf{u}(i)$ denotes the ground-truth MVF. We can see in the figure that MHBA shows the best performance among the five methods while IBME shows the worst. For the test videos *Coastguard* and *City*, SHBA-BF shows performances worse than SHBA even if it utilizes duality of MVF during simultaneous estimation. This is because the parameters ζ and l, which control the contribution of the dual MV to the target MV during refinement process, are not fine-tuned for each test video.



(a) *City* (b) *Mobile*

(c) *Ice* (d) *Silent*

Figure 5.17: PSNR performance comparisons.

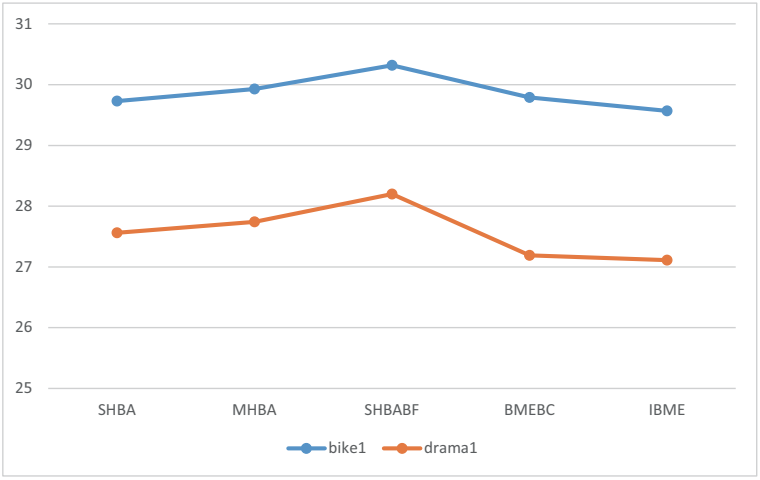


Figure 5.18: Average PSNR performance for the videos containing objects with fast and dynamic motion characteristics.

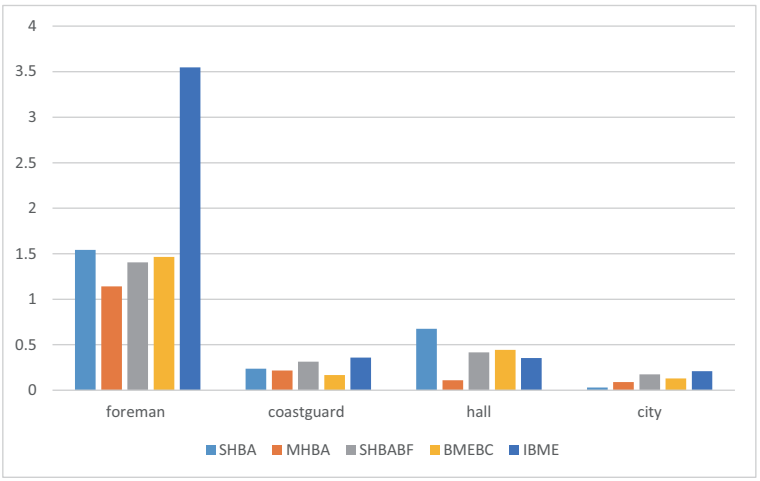


Figure 5.19: Average estimation error performance.

Table 5.1: Average PSNR(dB) Performances of Proposed Methods for CIF Test Videos

Sequences	SHBA	MHBA	SHBA-BF
Foreman	31.74	31.78	31.83
Coastguard	32.60	32.60	32.58
Silent	33.76	33.79	33.86
Mobile	26.63	26.83	26.70
Hall	32.70	32.72	32.77
City	33.45	33.61	33.50
Highway	30.93	30.96	30.99
Harbour	29.16	29.19	29.25
Ice	30.91	31.11	31.19
Flower	27.97	28.00	28.01
Average	30.98	31.05	31.06

Table 5.2: Average SSIM Performances of Proposed Methods for CIF Test Videos

Sequences	SHBA	MHBA	SHBA-BF
Foreman	0.9327	0.9333	0.9333
Coastguard	0.9480	0.9480	0.9476
Silent	0.9572	0.9576	0.9579
Mobile	0.9578	0.9601	0.9585
Hall	0.9594	0.9602	0.9601
City	0.9496	0.9516	0.9499
Highway	0.9318	0.9317	0.9328
Harbour	0.9301	0.9303	0.9312
Ice	0.9662	0.9679	0.9681
Flower	0.9625	0.9629	0.9630
Average	0.9495	0.9504	0.9502

Table 5.3: Average PSNR(dB) Performances of Proposed Methods for qFHD Test Videos

Sequences	SHBA	MHBA	SHBA-BF
Walkinglion	34.76	34.91	34.88
Soccer	31.19	31.54	31.40
Drama2	32.54	32.58	32.67
Runtrack	33.08	33.53	33.12
Subtitle	35.36	35.46	35.48
Bike1	29.26	29.44	29.39
Bike2	35.96	36.09	36.04
Drama1	27.95	28.20	28.38
Diving	29.95	30.05	29.99
Biketrack	33.95	34.25	34.04
Average	32.40	32.60	32.54

Table 5.4: Average SSIM Performances of Proposed Methods for qFHD Test Videos

Sequences	SHBA	MHBA	SHBA-BF
Walkinglion	0.9633	0.9648	0.9645
Soccer	0.9781	0.9805	0.9792
Drama2	0.9788	0.9791	0.9793
Runtrack	0.9811	0.9826	0.9816
Subtitle	0.9949	0.9952	0.9951
Bike1	0.9317	0.9344	0.9332
Bike2	0.9843	0.9849	0.9846
Drama1	0.9517	0.9532	0.9543
Diving	0.9488	0.9502	0.9492
Biketrack	0.9849	0.9863	0.9855
Average	0.9697	0.9711	0.9707

Table 5.5: Average PSNR (dB) Performance, with Constant and Variable δ

	constant δ		variable δ
Video	δ	PSNR	PSNR
Foreman	768	31.68	31.74
Coastguard	256	32.53	32.60
Mobile	896	26.59	26.63
City	576	33.48	33.45
Ice	640	30.56	30.91

Table 5.6: Average PSNR(dB) Performances of Existing Methods for CIF Test Videos

Sequences	VRP (k=6)	EVR (k=6)	CVR (k=6)	CMMR	MFMAP (k=3)	PRBME	BMEBC	HBMERM	IterME
Foreman	30.93	30.84	31.36	31.22	31.71	31.62	31.83	31.77	31.47
Coastguard	28.87	29.58	29.85	31.31	31.77	31.66	32.48	32.32	31.93
Silent	32.83	33.13	33.37	33.25	33.57	33.50	33.86	33.64	33.35
Mobile	25.02	24.50	24.89	25.30	26.34	25.79	26.84	26.28	26.69
Hall	31.78	32.46	32.44	32.28	32.71	32.55	32.75	32.65	32.52
City	32.72	32.08	32.31	31.76	33.43	32.69	33.51	33.35	33.25
Highway	29.11	29.77	30.29	30.30	31.04	31.11	31.00	31.21	30.98
Harbour	27.54	28.16	28.21	28.37	28.91	28.52	29.16	28.91	28.70
Ice	28.27	28.85	29.72	29.41	31.14	30.24	31.14	30.11	29.07
Flower	27.21	27.30	27.61	26.92	27.97	27.82	28.02	27.91	27.80
Average	29.42	29.66	30.00	30.01	30.85	30.55	31.05	30.81	30.57

Table 5.7: Average SSIM Performances of Existing Methods for CIF Test Videos

Sequences	VRP (k=6)	EVR (k=6)	CVR (k=6)	CMMR	MFMAP (k=3)	PRBME	BMEBC	HBMERM	IterME
Foreman	0.9178	0.9173	0.9238	0.9285	0.9301	0.9317	0.9336	0.9366	0.9313
Coastguard	0.8529	0.8697	0.8730	0.9358	0.9302	0.9353	0.9469	0.9459	0.9428
Silent	0.9430	0.9488	0.9480	0.9518	0.9551	0.9548	0.9592	0.9565	0.9560
Mobile	0.9295	0.9161	0.9261	0.9340	0.9549	0.9493	0.9608	0.9560	0.9582
Hall	0.9472	0.9511	0.9557	0.9560	0.9596	0.9590	0.9604	0.9597	0.9592
City	0.9412	0.9340	0.9364	0.9535	0.9492	0.9413	0.9503	0.9490	0.9469
Highway	0.9155	0.9164	0.9240	0.9206	0.9319	0.9366	0.9338	0.9353	0.9317
Harbour	0.9039	0.9138	0.9153	0.9182	0.9268	0.9225	0.9299	0.9279	0.9245
Ice	0.9382	0.9466	0.9578	0.9303	0.9679	0.9657	0.9688	0.9656	0.9524
Flower	0.9475	0.9476	0.9546	0.9502	0.9627	0.9603	0.9632	0.9613	0.9604
Average	0.9237	0.9261	0.9315	0.9379	0.9468	0.9457	0.9507	0.9494	0.9463

Table 5.8: Average PSNR(dB) Performances of Existing Methods for qFHD Test Videos

Sequences	VRP (k=6)	EVR (k=6)	CVR (k=6)	CMMR	MFMAP (k=3)	PRBME	BMEBC	HBMERM	IterME
Walkinglion	30.12	32.75	33.37	33.51	34.46	34.23	34.73	34.61	34.42
Soccer	27.60	26.89	27.25	29.65	29.97	30.76	30.61	30.93	30.41
Drama2	30.02	29.75	29.98	30.65	31.67	30.85	32.57	31.54	31.27
Runtrack	27.05	27.24	27.48	31.44	31.79	33.12	32.68	33.42	32.95
Subtitle	28.81	32.08	34.06	33.72	35.10	35.02	35.38	35.25	34.75
Bike1	26.31	26.86	28.07	28.60	28.89	28.81	29.48	29.17	29.20
Bike2	32.33	31.57	33.18	34.49	35.74	35.44	35.88	35.70	35.17
Drama1	26.31	25.93	26.21	27.14	27.44	27.21	27.81	27.54	27.54
Diving	28.84	28.88	29.33	29.63	29.57	29.67	29.97	29.81	29.75
Biketrack	30.64	29.99	30.85	32.44	32.91	32.66	33.34	33.38	33.57
Average	28.82	29.19	29.96	31.12	31.75	31.77	32.24	32.13	31.90

Table 5.9: Average SSIM Performances of Existing Methods for qFHD Test Videos

Sequences	VRP (k=6)	EVR (k=6)	CVR (k=6)	CMMR	MFMAP (k=3)	PRBME	BMEBC	HBMERM	IterME
Walkinglion	0.9321	0.9349	0.9410	0.9560	0.9600	0.9598	0.9614	0.9638	0.9617
Soccer	0.9344	0.9169	0.9237	0.9648	0.9684	0.9755	0.9738	0.9776	0.9746
Drama2	0.9648	0.9617	0.9639	0.9706	0.9758	0.9738	0.9791	0.9756	0.9721
Runtrack	0.9195	0.9218	0.9271	0.9750	0.9724	0.9830	0.9780	0.9840	0.9814
Subtitle	0.9789	0.9877	0.9927	0.9922	0.9943	0.9944	0.9947	0.9949	0.9936
Bike1	0.8995	0.8970	0.9198	0.9290	0.9278	0.9274	0.9322	0.9324	0.9287
Bike2	0.9678	0.9546	0.9707	0.9792	0.9834	0.9828	0.9837	0.9834	0.9808
Drama1	0.9226	0.9201	0.9292	0.9416	0.9466	0.9459	0.9511	0.9468	0.9431
Diving	0.9346	0.9296	0.9400	0.9461	0.9456	0.9453	0.9483	0.9476	0.9480
Biketrack	0.9646	0.9576	0.9675	0.9801	0.9801	0.9807	0.9825	0.9840	0.9845
Average	0.9418	0.9381	0.9475	0.9634	0.9654	0.9668	0.9684	0.9690	0.9668

Table 5.10: Average Computing Time (seconds per frame) for CIF Videos

Algorithm	times (s)
VRP (k=6)	1.108
EVR (k=6)	1.842
CVR (k=6)	1.636
CMMR	2.02
MFMAP (k=3)	987.3
PRBME	41.7
BMEBC (k=3)	over 1000
IterME	over 1000
Proposed (k=6)	1.43

Chapter 6

Conclusion

In this dissertation, three approaches to motion vector field (MVF) refinement were proposed for block-based motion-compensated frame interpolation (MCFI). In chapter 2, the first approach, called *single hypothesis Bayesian approach* (SHBA), was proposed to estimate the true MVF from its observed MVF by maximizing the posterior probability of it. The observation is assumed to be a degraded version of the true MVF, and the degradation was modeled as locally stationary additive Gaussian noise (AGN). The prior distribution of the true MVF is designed to rely on the distances between the MV and its neighbors and to properly smooth false MVs in the observation. The maximum *a posteriori* (MAP) solution was obtained via iterative conditional mode (ICM) method. The result is an iterative update equation, which produces the kth estimate of the true MV of a block by combining, according to the estimated noise variance, the observation and the neighboring $(k-1)$th estimates, the latter themselves having been averaged with varying weights. The experimental results in chapter 5 showed that, owing to the proposed noise variance estimation method and the proposed prior distribution model, SHBA achieves performance comparable to or better than the state-of-the-art BME algorithms [21][22][33]. Nevertheless, SHBA, as a MR technique, takes

far less computation as compared to the algorithms. Therefore, this dissertation proves that a reasonable observation MVF contains virtually all the candidates for the best estimate of the true MVs.

In chapter 3, the assumption was extended into that multiple observed MVFs, which are the results of a BME incorporating multiple block sizes for matching, are degraded versions of the true MVF. The true was then estimated from the multiple observations by maximizing the posterior probability of it. The MAP solution was obtained via ICM method. The result is an iterative update equation, which produces the kth estimate of the true MV of a block by combining, according to the estimated noise variances, the multiple observations and the neighboring $(k-1)$th estimates, the latter themselves having been averaged with varying weights. In addition, the noise variance adjustment method was proposed in order to solve motion boundary problem. The estimated noise variances are adaptively adjusted according to how suitable the corresponding observation MV is for the target block. Experimental results in chapter 5 showed that the proposed approach, called *multiple hypotheses Bayesian approach* (MHBA), outperforms not only SHBA but also the state-of-the-art BME algorithms. It showed good performance especially on the videos with periodic patterned areas.

Finally, in chapter 4, the third approach, called *single hypothesis Bayesian approach in a bidirectional framework* (SHBA-BF) was proposed to estimate the true forward and backward MVFs simultaneously. The true MVFs were obtained by maximizing the joint posterior probability of them from the observed forward and backward MVFs under the assumption that the observations are degraded versions of the true by locally stationary AGN. The joint prior distribution model was proposed in order to fully utilize duality of MVF under simultaneous estimation. The MAP solution was obtained by the alternation method via ICM. Experimental results in chapter 5 showed that SHBA-BF outperforms not only SHBA but also the existing methods. It showed good performance especially on the videos with objects having dynamic motions.

Bibliography

[1] E. Dubois, "Motion-compensated filtering of time-varying images," *Multidimensional Systems and Signal Processing*, Vol. 3, pp. 211-239, 1992.

[2] H. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, pp. 185-203, 1981.

[3] H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 5, pp. 565-593, September 1986.

[4] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 9, pp. 910-927, September 1992.

[5] C. Stiller, "Object-based estimation of dense motion field," *IEEE Transactions on Image Processing*, Vol. 6, No. 2, pp. 234-250, February 1997.

[6] M. M. Chang, A. M. Tekalp, and M. I. Sezan, "Simultaneous motion estimation and segmentation," *IEEE Transactions on Image Processing*, Vol. 6, No. 9, pp. 1326-1333, September 1997.

[7] M. J. Black, P. Anandan, "A model for the detection of motion over time," in *Proceedings of ICCV 1990*, pp. 33-37, 1990.

[8] M. J. Black, "The robust estimation of multiple motions: parametric and piecewise-smooth flow field," *Computer Vision and Image Understanding*, Vol. 63, No. 1, pp. 75-104, 1996.

[9] J. Wei, Z.-N. Li, "An efficient two-pass MAP-MRF algorithm for motion estimation based on mean field theory," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 6, pp. 960-792, September 1999.

[10] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-17, No. 6, pp.576-586, June 1995.

[11] K. P. Lim, M. N. Chong, and A. Das, "Low-bit-rate video coding using dense motion field and uncovered background prediction," *IEEE Transactions on Image Processing*, Vol. 10, No. 1, pp. 164-166, January 2001.

[12] K. P. Lim, A. Das, and M. N. Chong, "Estimation of occlusion and dense motion fields in a bidirectional bayesian framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 712-718, May 2002.

[13] S. Ince, and J. Konrad, "Occlusion-aware optical flow estimation," *IEEE Transactions on Image Processing*, Vol. 17, No. 8, pp. 1443-1451, August 2008.

[14] F. Xu, and Q. Dai, "Occlusion-aware motion layer extraction under large interframe motions," *IEEE Transactions on Image Processing*, Vol. 20, No. 9, pp. 2615-2626, September 2011.

[15] A. Ayvaci, M. Raptis, and S. Soatto, "Sparse occlusion detection with optical flow," *International Journal of Computer Vision*, October 2011.

[16] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi, "Bilateral filtering-based optical flow estimation with occlusion detection," in *Proceeding of the European Conference on Computer Vision*, 2006.

[17] R. Ben-Ari, and N. Sochen, "Variational stereo vision with sharp discontinuities and occlusion handling," in *Proceeding of ICCV*, 2007.

[18] R. Depommier, and E. Dubois, "Motion estimation with detection of occlusion areas," in *Proceeding of ICASSP*, 1992.

[19] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," in *Proceeding of CVRP*, 2005.

[20] G. Haan, P. W. A. C. Biezen, H. Huijgen, and O. A. Ojo, "True-motion estimation with 3-D recursive search block matching," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 3, No. 5, pp. 368-379, October 1993.

[21] D. Wang, L. Zhang, and A. Vincent, "Motion-compensated frame rate up-conversion - Part I: Fast multi-frame motion estimation," *IEEE Transactions on Broadcasting*, Vol. 56, No. 2, pp. 133-141, June 2010.

[22] M. Santoro, G. Alregib, and Y. Altunbasak, "Joint framework for motion validity and estimation using block overlap," *IEEE Transactions on Image Processing*, Vol. 22, No. 4, pp. 1610-1619, April 2013.

[23] R. Thoma, and M. Bierling, "Motion compensating interpolation considering covered and uncovered background," *Signal Processing: Image Communication*, Vol. 1, Issue. 2, pp. 191-212, October 1989.

[24] G. Koo, K. W. Lim, and S. J. Choi, "Complementary block-based motion estimation for frame rate up-conversion," in *Proceeding of ICCE*, 2011.

[25] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory of warping," in *Proceeding of European Conference on Computer Vision*, 2004.

[26] M. Proesmans, L. B. Gool, E. Pauwels, and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion," in *Proceeding of European Conference on Computer Vision*, 1994.

[27] S. Ince, and J. Konrad, "Geometry-based estimation of occlusions from video frame pairs," in *Proceeding of ICASSP*, March 2005.

[28] K.-S. Choi and S.-J. Ko, "Hierarchical motion estimation algorithm using reliable motion adaption," *Electronics Letters*, Vol. 46, No. 12, June 2010.

[29] B.-D. Choi, J.-W. Han, C.-S. Kim and S.-J. Ko, "Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 4, pp.407-416, April 2007.

[30] K. S. Seo, and J. K. Kim, "Hierarchical block-motion estimation using liear model-based postprocessing," *Electronics Letters*, Vol. 29, No. 22, pp. 1915-1917, October 1993.

[31] H. Liu, R. Xiong, D. Zhao, S. Ma, and W. Gao, "Multiple hypotheses bayesian frame rate up-conversion by adaptive fusion of motion-compendated interpolations," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 8, August 2012.

[32] S.-C. Tai, Y.-R. Chen, Z.-B. Huang and C.-C. Wang, "A multi-pass true motion estimation scheme with motion vector propagation for frame rate up-conversion applications," *Journal of Display Technology*, Vol. 4, No. 2, pp.188-197, June 2008.

[33] D. Kim, H. Lim, and H. Park, "Iterative true motion estimation for motion-compensated frame interpolation," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 23, No. 3, March 2013.

[34] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proceedings of the IEEE*, Vol. 78, No. 4, pp. 678-689, April 1990.

[35] L. Alparone, M. Barni, F. Bartolini, and V. Cappellini "Adaptively weighted Vector-median filters for motion fields smoothing," *Proceedings of ICASSP*, Vol. 4, pp. 678-689, April 1990.

[36] S. Sekiguchi, Y. Idehara, K. Sugimoto, and K. Asai "A low-cost video frame-rate up conversion using compressed-domain information," *Proceedings of ICIP*, Vol. 2, 2005.

[37] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, Vol. 2, pp. 283-310, 1989.

[38] A. Singh, "Image-flow computation: an estimation-theoretic framework and a unified perspective," *CVGIP: Image Understanding*, Vol. 56, Issue 2, pp. 152-177, September 1992.

[39] H. Zheng and S. D. Blostein, "An error-weighted regularization algorithm for image motion-field estimation," *IEEE Transactions on Image Processing*, Vol. 2, No. 2, pp. 246-252, April 1993.

[40] Y. W. Sohn, and M. G. Kang, "Block-based motion vector smoothing for periodic pattern region," *ICIAR 2007*, LNCS 4633, pp. 491 500, 2007.

[41] S.-H. Lee, O. Kwon, and R.-H. Park, "Motion vector correction based on the pattern-like image analysis," *IEEE Transactions on Consumer Electronics*, Vol. 49, No. 3, pp.479-484, August 2003.

[42] A.-M. Huang, and T. Q. Nguyen, "A multistage motion vector processing method for motion-compensated frame interpolation," *IEEE Transactions on Image Processing*, Vol. 17, No. 5, pp. 694-708, May 2008.

[43] A.-M. Huang, and T. Q. Nguyen, "Correlation-based motion vector processing with adaptive interpolation scheme for motion-compensated frame interpolation," *IEEE Transactions on Image Processing*, Vol. 18, No. 4, pp. 740-752, April 2009.

[44] G. Dane, and T. Q. Nguyen, "Motion vector processing for frame rate up conversion," *ICASSP 2004*, Vol. 3, pp.309-312, 2004.

[45] N. Jacobson, Y.-L. Lee, V. Mahadevan, N. Vasconselos, and T. Q. Nguyen, "A novel approach to FRUC using discriminant saliency and frame segmentation," *IEEE Transactions on Image Processing*, Vol. 19, No. 11, pp. 2924-2934, November 2010.

[46] B. Girod, "The efficienty of motion-compensating prediction for hybrid coding of video sequences," *IEEE Journal on Selected Areas in Communications*, Vol. 5, Issue. 7, pp. 1140-1154, August 1987.

[47] J. Besag "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, Series B 48, pp. 259-302, 1986.

[48] I. Patras, E. A. Hendriks, and R. L. Lagendijk, "Probabilistic confidence measures for block matching motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 8, pp. 988-995, August 2007.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality accessment: From error visibility to structual similarity," *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp.1-14, April 2004.

[50] G. Dane and T. Q. Nguyen, "Optimal temporal interpolation filter for motion-compensated frame rate up conversion," *IEEE Transactions on Image Processing*, Vol. 15, No. 4, pp.978-991, April 2006.

[51] A. M. Zoubir, V. Kolvunen, Y. Chakhchoukh, and M. Muma, “Robust estimation in signal processing,” *IEEE Signal Processing Magazine*, Vol. 29, No. 4, pp. 61-80, July 2012.

[52] A. Blake, P. Zisserman, “Visual Reconstruction,” Cambridge, MA: MIT Press.

[53] Stan Z. Li, “Markov random field modeling in image analysis,” Springer, 2009.

국문 초록

본 논문은 움직임 보상 프레임 보간을 위한 최대 사후 확률 기반의 움직임 벡터장 교정 방법을 제안한다. 먼저 single Hypothesis Bayesian approach (SHBA)는 블록 기반의 움직임 예측 결과물인 관찰 움직임 벡터장으로부터 실제 움직임 벡터장을 예측하기 위하며 제안되었다. 이를 위하여 관찰 움직임 벡터가 실제 움직임 벡터의 가우시안 열화를 통하여 얻어졌다 가정하며 관찰 움직임 벡터의 신뢰도를 반영하는 열화 잡음의 분산을 이웃하는 관찰 움직임 벡터들로부터 직접 예측한다. 또한 실제 움직임 벡터장의 사전 분포를 이웃하는 움직임 벡터들 사이의 닮음에 적응적으로 의존하도록 설계하였다. 다음으로 multiple Hypotheses Bayesian approach (MHBA)는 다수의 관찰 움직임 벡터장들로부터 실제 움직임 벡터장을 예측하기 위하여 제안되었으며 이 때 다수의 관찰 움직임 벡터장은 서로 다른 크기의 블록을 이용한 블록 기반의 움직임 예측 결과라 가정한다. 다수의 관찰 벡터들이 실제 움직임 벡터의 가우시안 열화로부터 얻어졌다 가정하며 관찰 벡터들로 부터 예측된 분산은 움직임 경계문제의 해결을 위하여 적응적으로 변경된다. 마지막으로 SHBA in a bidirectional framework (SHBA-BF)은 SHBA를 바탕으로 실제 정방향 및 역방향 움직임 벡터장을 동시에 예측하기 위하여 제안되었다. 관찰 과정을 위하여 가우시안 열과를 가정하였으며 움직임 벡터장의 이중성을 이용하기 위하여 정방향 및 역방향 움직임 벡터장의 공동 사전 분포를 새롭게 제안하였다.

주요어: 움직임 교정, 실제 움직임 예측, 움직임 보상 프레임 보간

학번: 2008-30246

감사의 글

사랑하는 부모님 최병희, 임순애 님께 이 논문을 바칩니다.